SHOULD ELIMINATIVE MATERIALISM BE ELIMINATED?

John Gordon, B.A.

Submitted in partial fulfilment of
the requirements for the degree of
DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF EDINBURGH

September 1997

DECLARATION

This thesis has been composed by the candidate, and is the candidate's own work. Parts of the thesis were discussed, in draft form, with Dr D Walsh, Dr G Madell, and Mr S Priest of the University of Edinburgh Department of Philosophy.

No part of this work has been submitted for any other degree, diploma or professional qualification, nor has any part of the thesis been published prior to submission for the award of PhD.

Signed:

29th January 1998

ABSTRACT

The thesis consists of a critical evaluation of Paul Churchland's eliminative materialism. The first of the central claims of the thesis is that it is unclear how radical an eliminativism Churchland wishes to adopt, in that his published work appears to vacillate between a position which is too modest to be regarded as a genuinely eliminative form of materialism, and positions which, while radical in their eliminativism, are not supported by the empirical evidence which Churchland presents. I conclude that eliminative materialism is itself a candidate for elimination, on the grounds that the radical positions are insupportable in principle, while the only eliminativist position which is defensible effectively fails to qualify as a form of 'eliminative materialism'.

The early chapters of the thesis consider the negative element in Churchland's position - the claim that 'folk psychology' will not ultimately reduce to neuroscience, and that, this being a constraint on the acceptability of folk psychology as a putative source of mental explanation, folk psychology ought therefore to be eliminated. The positive element of Churchland's position is then considered: his claims that research programmes currently being undertaken in both parallel distributive processing ('PDP'), and neuroscience, converge in providing a more psychologically realistic account of human cognition than do more conventional accounts, which utilise the explanatory categories of folk psychology; and that, as this PDP model eschews the use of folk psychological categories, it thus entails elimination of folk psychology, as anticipated in the earlier, negative thesis. I consider the philosophical implications of

the PDP model, and find it deficient with regard to two of the main areas of philosophical interest considered by Churchland: the operation of moral choice; and consciousness. My conclusion is that PDP does not serve to provide support for any but the most modest of eliminativist positions - so that what empirical plausibility PDP may have cannot rescue eliminative materialism from elimination.

CONTENTS

Word count: 68428

CHAPTER 1 INTERTHEORETIC REDUCTION AND ELIMINATIVISM

Paul Churchland cites, as the distinguishing feature of eliminative materialism: '... its denial that a smooth intertheoretic reduction is to be expected ... of the framework of folk psychology to the framework of a matured neuroscience'.[1] It is, I will later argue, often unclear precisely what eliminativists propose to eliminate. Here, however, the target is clear: explanation of human action in terms of folk psychology. The eliminativist conclusion requires the addition of a prior premise:

i. Folk psychology provides genuine explanation of human action iff it is smoothly reducible to mature neuroscience;

ii. Folk psychology is irreducible to neuroscience; therefore

iii. Folk psychology does not provide genuine explanation of human action.

If we take the term 'folk psychology' to denote explanation in terms of propositional attitudes such as beliefs and desires, then what is being denied by Churchland is the possibility of genuine explanation which utilises beliefs and desires among its explanatory categories.[2]

As the argument is valid, the truth of the conclusion depends upon the truth of the premises.

The second premise is widely held to be true: clearly the eliminativist requires the truth of this premise, but non-reductive physicalists such as Fodor and Davidson also accept the truth of the second premise. Their case for rejecting the eliminativist conclusion thus depends upon proving the falsity of the first premise. Dualists will also accept the falsity of the first premise, and the truth of the second premise. I propose in this chapter to consider the controversial first premise, returning later to the question of whether folk psychology is reducible.

I take it that for Churchland, it is necessary for (m) being an explanation of (p) that (m) be a cause of (p). Where (m) has no causal efficacy, explanations which appeal to the presence or absence of (m) may have instrumental value, but they are not 'genuine' explanations. The first premise thus makes it a constraint on (m) being genuinely explanatory that (m) be smoothly reducible to some explanatory category of mature neuroscience. 'Maturity' would reside in the completeness of neuroscience: it would be unrevisable, in the sense of the unrevisability aimed for by Descartes; there would be no explanatory gaps, no events within its domain which could not be predicted, given sufficient knowledge of antecedent events, and no errors in the theory. Both Paul and Patricia Churchland emphasise the continuing evolution of the sciences, so that what is being considered here is a future version of neuroscience. 'Smooth' reduction would be type reduction, where the explanatory categories of folk psychology are shown to be identical to those of neuroscience. This must be distinguished from a weaker form of reduction, where every event is shown to be a physical event. If a physical event can have mental properties, then if we assume that it is in virtue of its properties that an event has causal efficacy, this weaker form of reduction - token event physicalism -

would leave mentalistic ('folk psychological') explanation autonomous. By 'autonomous' I mean that folk psychology would generate explanation which was beyond the explanatory reach of neuroscience. It is this autonomy which Churchland is concerned to deny - hence the eliminativist's constraint that for folk psychological explanation to be genuine, there must be reduction in the stronger form of type-reduction.[3]

Lennon and Charles see such reduction as having the epistemic advantage of clarifying or simplifying our understanding:

> Reductionist accounts aim to show that where we thought we had two sets of concepts, entities, laws, explanations, or properties, we in fact have only one, which is most perspicuously characterised in terms of the reducing vocabulary.[4]

Lennon and Charles provide five conditions for reduction:

> (1) the derivation of higher-level laws, (2) the discovery of nomological biconditionals linking the terms of each theory, (3) the presence of genuine properties at the reducing level, (4) the causal explanation in terms of the reducing theory of the phenomena explained by the reduced theory, and (5) a reason for giving privileged status to the reducing descriptions.[5]

Condition (1) presents an immediate problem: an exhaustively deterministic and nomological universe is a metaphysical assumption of scientistically-minded philosophers such as the Churchlands.[6] If we adhere to the view that human beings have free will - that they can resist acting on the rational patterns of behaviour to which their beliefs and desires would appear to commit them - then this entails that there can be no laws governing human action.

Condition (2) concerns 'bridge laws'. It appears that what is required is that there be a law of nature such that, for example, each occurrence of a given mental property is identical to some physical (or, in the present case, neurological) property. The claim that some mental property is identical with a neurological property is an empirical claim. It is difficult to see what would count as evidence in support of the claim. Suppose the co-occurrence of some mental property and some neurological property is established by frequent testing for the presence of each property. This would be consistent with identity (what is being observed is the same entity, variously described), but co-occurrence is also consistent with the truth of dualism for both entities and properties. Nor can I see how the fact of this relationship being covered by a law of nature could be demonstrated. In short, all that is likely to be available to the reductivist is an accumulation of inductive evidence, which must then be interpreted in the light of his own metaphysical assumptions.

Condition (3) alludes to a point which I made earlier: that for a property to be 'genuine' in the sense required is for it to have causal efficacy. This entails that, qua physical causal property, it will be denoted by a predicate which will form the antecedent in some physical law. Thus, for example, the hard disk in my computer is a physical object which has various properties. Some of these properties, such as the property of being currently located in Edinburgh, is not a genuine property in the requisite sense.

Condition (4) - the explanation of the reduced by the reducing theory - is, as earlier suggested, a motivation for reduction. The promise would thus be that matured neuroscience would be capable of explaining and predicting any events hitherto explained or predicted in terms of beliefs and desires. Neuroscience would also be able to explain how folk psychology had what explanatory and predictive success it had had, and how and why it was deficient. In addition to the enrichment of our understanding, reduction serves to provide vindication for the successfully reduced and successfully reducing theories. The reducing theory is privileged - condition (5): as the reduction relation is asymmetrical, the superiority of neuroscience to folk psychology would be guaranteed on epistemic grounds.

In addition to the epistemological motivation for reduction, there is an ontological motivation. This is made explicit by Lennon and Charles, who provide as a motivation that:

> ... many would accept (1) that social and psychological events are (in some sense) constituted out of physical events, (2) that physical explanation is complete, i.e. that all physically characterizable events are susceptible to explanation in terms of physically sufficient causes, and (3) that there are causal explanations employing social and psychological vocabulary.[7]

The fact that 'many would accept' these propositions does not entail their truth; the first two are a posteriori and contingent and, I suggest, false. To make an ontological case for reduction of the psychological to the physical on the basis of the first of the propositions - which is a definition of physicalism - would be question-begging. This ontological assumption appears to underlie the epistemological case for reduction,

however: there are no grounds for expecting that some physical reducing theory will, on achieving maturity, succeed in explaining all events in the domain of folk psychology, other than the assumption that physicalism is true. The second of the propositions - the completeness of physical explanation - is a sufficient condition for physicalism. If it is true that there are physical sufficient causes for all physical events, then either all mental events are physical events, in which case token physicalism is true, or there are no mental events. A further possibility is epiphenomenalism regarding the mental. Epiphenomenalism is consistent with every cause being a physical cause, but is inconsistent with the third of Charles and Lennon's conditions for reduction: the presence of genuine properties at the reducing level, as mental properties would on the epiphenomenalist account be causally inert.[8] One final possibility which is consistent with the second proposition is that there are both physical sufficient causes and mental sufficient causes for (at least some) physical events. This possibility would be inconsistent with two apparent assumptions of the physicalist reductionist: the ontological primacy of the physical[9] - if mental causes are sufficient for physical events, then physical causes are not necessary for these events (if (a) is sufficient for (c), then (b) is not necessary for (c)). ; and the need for parsimony in explanation - the latter being explicit in the reductionist's desire to demonstrate that where we had thought there were two explanations, there is in fact only one. Charles and Lennon's third proposition - that some causal explanation is couched in psychological terms - is uncontroversially true - though it is consistent with instrumentalism regarding the mental, which is in turn inconsistent with the reductionist requirement of genuine properties at the reducing level.

Patricia Churchland and Sejnowski offer a pragmatic motivation for seeking to reduce: the process of reducing - or attempting to reduce - will lead to new scientific discovery, and hence take us further along the road to matured science:

> ... seeking reductions of macro-level theory to micro-level theory does not imply that one must first know everything about the elements of the micro theory before research at the macro level can be usefully undertaken. Quite the reverse is advocated - research should proceed at all levels of the system, and *co-evolution of theory may enhance progress at all levels.*(my emphasis)[10]

The apparent failure of folk psychology to contribute to the evolution of neuroscience (or vice-versa) is thus taken as strong prima facie evidence for the falsity of folk psychology, presumably on the grounds that if reducibility entails co-evolution, then the absence of co-evolution entails the absence of reducibility. Nor can folk psychology, on this account, be safe from the requirements of reducibility until such time as neuroscience is fully matured: the co-evolution thesis recognises the current shortcomings of neuroscience, and looks to psychology to contribute to their resolution:

> Neuroscience and psychology need each other. Crudely, neuroscience needs psychology because it needs to know what the system does ... psychology needs neuroscience for the same reason ...[11]

This claim is tenable only in terms of Churchland's sparse account of cognition: as the processing of inputs and outputs. If one suggests that 'what the system does' is to represent external reality to itself via the manipulation of propositional attitudes which have semantic content, to distinguish between self and other, or to experience qualia,

then it seems that neuroscience will make little if any progress in providing a 'micro-level' explanation of these events. Here again, the advantage being claimed for reduction depends upon question-begging: the mental realm must be interpreted so as to eliminate in advance all that is characteristically (and irreducibly) mental, in order that there is the possibility of reduction. This cannot be taken to entail what the eliminativist wants: namely, that the irreducibility of mentalistic psychology entails its radical falsity.

Churchland and Sejnowski's concern to advance physical science reveals what is perhaps the most pressing motivation for intertheoretic reduction: the aim of the unity of science. This is a central objective of the philosopher, on Paul Churchland's account: '... philosophy remains the focal point for synoptic concerns about how the scattered sciences can be fitted together into a unified and coherent account of the world and our place in it.'[12] Churchland offers a further task for the philosopher, which again suggests a high degree of deference to physical science: the development of proto-scientific speculation into testable empirical theory. One conspicuous omission from Churchland's account of what philosophy is is conceptual analysis. He argues that this project, together with its target of the discovery of synthetic a priori truth, is doomed to failure.

The possibility in principle of the reduction of all scientific theories to the most basic scientific theory is a necessary and sufficient condition for the unity of science - the pursuit of which is a legacy of logical positivism. Sorell, in his aptly-titled

Scientism: Philosophy and the Infatuation with Science, quotes from Carnap: '... when we say that scientific knowledge is unlimited, we mean: there is no question whose answer is in principle unattainable by science'.[13] This is an exceptionally strong claim. It entails, for example, the principled possibility of an exhaustively deterministic and scientific account of every future event, so that, for example, both the outcome of the next general election, and the contents of the menu at 10 Downing Street on the day following the election, would be predictable by science. Notwithstanding Churchland's reductive account of philosophical method, it seems sufficient to claim that this is massively counterintuitive. What Carnap is here claiming is sufficient, but not necessary for the unity of science, and could perhaps be described as the 'Unity of all Knowledge' thesis.[14]

Sorell locates the origins of contemporary scientism in Carnap's 'scientific empiricism', which is based on five claims:

> 1) science is unified; (2) there are no limits to science; (3) science has been enormously successful at prediction, explanation and control; (4) the methods of science confer objectivity on scientific results; and (5) science has been beneficial for human beings.[15]

Clearly claim (1) is merely a statement of the unity of science. The claim that science is unified in this way is a metaphysical claim. It entails a further metaphysical claim: that monism is true. The integration of all explanation will necessitate the linking, via bridge laws, of all of the properties and laws of each special science. If those properties and laws are fundamentally distinct in one or more of the sciences, then this ontological pluralism will entail explanatory pluralism, and hence the falsity of the

claim that science is unified. There can be no a priori argument for the claim that monism is true, and hence reductionism and unity is possible.[16] The Carnapian unity of science claim, when combined with the further a posteriori and metaphysical claim that ultimately all that exists is physical, jointly entail the elimination of any irreducibly mental entities. Hence Stich suggests that:

> ... if our best theories fail to quantify over putative entities of a certain sort, we should conclude that there are no such things. And since (ex hypothesis) cognitive science does not invoke the language or concepts of folk psychology, the states of folk psychology are not among the entities over which it quantifies. So these putative states do not exist.[17]

If the unity of science thesis is false, then this entails the falsity of Stich's claim - hence the importance of the unity of science project for the eliminativist. The quotation also assumes what remains to be proved: that folk psychology is a theory, and hence in competition with or reducible to our 'best' theories. The term 'best theories' is tendentious. There are three possible grounds for arguing that physics is the 'best theory': as I have already suggested, one might argue that the best theory will be the one which, by virtue of its greatest explanatory and predictive reach, can successfully reduce other theories, while not itself being subject to reduction. Whether physics can achieve this vis a vis the mental depends in part on the truth of the physicalist's ontological claim; if the ontological claim is false, then this entails the impossibility of physical reduction of any sciences which incorporate mental realism. A second ground for holding that physics is the best theory might be the truth of physicalism: if as Charles and Lennon suggest, '...psychological events are ... constituted out of physical events', then the science which treats most directly of the physical would appear to be

the science which is closest to fundamental reality. As already urged, however, the physicalist's ontological claim is at best not proven. Finally, it could be argued that the superiority of physics resides in the fact that physics does not avail itself of the ceteris paribus clauses which are prevalent in the special sciences (and most particularly, the social sciences, where - as Charles and Lennon point out, mentalistic terminology is also most prevalent). The entailment from relative absence of such hedging to superiority of the theory is not clear - but is implicit in Paul Churchland's jibe that folk psychology is '... festooned with ceteris paribus clauses...'[18]

The first of Carnap's five claims - the unity of science - is thus especially significant for the eliminativist project - but the second and third of the claims are also at least implicit in contemporary eliminativism. The second claim is the already-discussed unity of all knowledge thesis. Although this thesis is much stronger than the eliminativist needs in order to eliminate folk psychology (assuming that the other eliminativist premises are true), the thesis does find a role in some of the more radical claims advanced in some of the writing of the Churchlands.

Carnap's third claim - the past success of science - is frequently cited by the Churchlands, where the success has been in terms of past reductions and eliminations of other theories by more basic physical theories. Thus, in addition to acceptance of the central metaphysical claims implicit in contemporary science, such as physicalism and determinism, an inductive argument from the history of science is appealed to in support of eliminativism. As already noted, Churchland sees the role of the philosopher as being proto-scientific and at least partly speculative. On occasion, this leads him to

a posture of relative open-mindedness with regard to the possibility of reducing folk psychology:

> How Computational Neuroscience and Connectionist AI will fare in the coming years remains to be seen ... whether folk psychological categories will find some kinematical and dynamical role within the new framework *remains a strictly open question.* (my emphasis)[19]

Carnap's third claim also informs Churchland's scepticism regarding folk psychology; he lambasts folk psychology as being 'stagnant' - having made little or no progress since the time of Homer. As with Stich's earlier-quoted claim that folk psychological entities do not exist, this stagnation argument contains the suppressed premise that folk psychology is a theory. If it is not, then the truth of the claim that science is unified does not entail the radical falsity of explanation in terms of propositional attitudes, and the eliminativist thesis - construed as the thesis that folk psychological explanation should be eliminated - falls. As the Unity of Science thesis is based on no more than induction from past reductive successes, it can only accumulate a progressively greater degree of probability as further successful reductions are achieved. As Searle notes, however, such successes are '... rather rare in the actual practice of science, and it is perhaps not surprising that the same half dozen examples are given over and over ...'.[20]

For Carnapian positivists, theories are sets of sentences, and explanation is via deduction from these sentences. This sentential model of theories will also facilitate intertheoretic reduction. A greatly simplified account of how this works is as follows:

the special science laws, once discovered, will be set out as sentences; the bridge laws will allow for their translation into sentences utilising the proprietary terminology of the reducing science, and once this has been carried out for all of the laws of the special science, reduction is achieved. In his essay 'On the Nature of Theories: A Neurocomputational Perspective', Churchland actually claims that the 'primary business of theories', on the conventional view, is 'prediction, explanation, and intertheoretic reduction'.[21] In this essay, Churchland rejects the sentential model of theories, on the grounds that this model makes the:

> ...fundamental assumption that language-like structures of some kind constitute the basic or most important form of representation in cognitive creatures, and the correlative assumption that cognition consists in the manipulation of those representations.[22]

Churchland's view is thus that elimination of propositional attitudes entails the elimination of the sentential model of theories.[23] The grounds for this view are, predictably, empirical, speculative, and based on research into brain function in order to establish how representation and computation *is* achieved in the absence of propositional attitudes.[24] Churchland's task is thus to give an alternative - 'neurophilosophical' account of what is a theory. The remainder of his essay fails to give a succinct answer. Churchland considers the purely physical relationships which the brain's billions of neurons may bear towards each other. Different configurations constitute different theories, each constituting a distinct paradigm, in the Kuhnian sense. Churchland makes the point that neuroscience is still at a very early stage, so that the promise is of a developing and progressively more sophisticated account of what is a theory.[25]

Churchland's account of theories is consistent with the eliminativist conclusion, but appears to cause problems for the eliminativist project, as it is inconsistent with some of his other central claims. Churchland needs, for example, to give necessary and sufficient conditions for being a theory, which will be consistent with both his neurophilosophical account, and his claim that folk psychology is a theory. In his 'Eliminative Materialism and the Propositional Attitudes', he cites the provision of ' ... explanations/ predictions/ understanding of ... behaviour' as criteria, under the section heading 'Why folk psychology is a theory'.[26] This is inconsistent with - and is explicitly contradicted by - Churchland in his 'On the Nature of Theories', in which this 'classical view' is dismissed: ' ... hardly anyone will now deny that there are serious problems with every element of the preceding picture.'[27] If, on the other hand, Churchland's putative necessary and sufficient conditions for being a theory are neurological conditions, then this will be inconsistent with the claim that folk psychology is a theory. If folk psychology is not a theory, then its failure to reduce to neuroscience theory is not a constraint on its acceptability.

It is, in any case, difficult to see how intertheoretic reduction can proceed in the absence of sentences, or some alternative modes of presentation which fulfil the same functions as sentences. This is because the object of reduction is to explain the laws of one theory in terms of another theory, and the possibility of achieving this cross-theoretical explanation is demonstrated via deduction. If it is a necessary condition for reduction that theories be expressible - and I take it that deduction and explanation do make this a necessary condition, then Churchland's account of what a

theory is renders reduction practically impossible, thereby making his claim for the irreducibility of folk psychology trivially true. The same objection would appear to rule out much less controversial reductions, such as that of meteorology to physics. An alternative non-Carnapian (i.e. non-sentential) model would be necessary. Churchland's position on theories thus appears to be incoherent: he needs a model of scientific unification which is both non-sentential and preserves the claim that folk psychology is a theory, but the claim that folk psychology is a theory cannot be separated from the claim that theories are sets of sentences.

An Ontological Version of Eliminativism.

The form of eliminativism considered up to this point advocates the elimination of a putative *theory* - folk psychology. On some construals, however, eliminative materialism also makes an ontological claim: the claim that there *are no* beliefs and desires; they do not exist. Thus Paul Churchland opens his essay 'Eliminative Materialism and the Propositional Attitudes' by defining eliminativism as '... the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so defective that *both the principles and the ontology* of that theory will eventually be displaced'.(my emphasis)[28]

The claim that there are no beliefs is a stronger claim that the claim that folk psychology is radically false. By 'stronger than', I mean that the former claim entails

the latter claim, but that the latter claim does not entail the former. Even if the proposition that folk psychology is radically false is true, a further premise is needed to entail the conclusion that there are no beliefs. One could, by analogy, argue that musical theory is in principle irreducible to more basic theory; that in addition to the acoustical properties of a concerto, there are meaningful and true propositions which may be put regarding concertos, and which pick out properties of concertos which cannot be explained by more basic theory. Notwithstanding this, concertos surely exist.[29] The missing premises, with regard to concertos, might be provided by a combination of what I earlier described as the 'Unity of all Knowledge' thesis, together with the claim that 'the entities we are committed to are simply those quantified over in our best scientific theories'.[30] The elimination of the concerto would then proceed as follows:

i. a smooth intertheoretic reduction of all theories to physics is a necessary and sufficient condition for the real existence of the entities quantified over by the theories which are candidates for reduction;

ii. musical theories of concertos are not so reducible; therefore

iii. concertos do not exist.

If we assume the truth of (ii), the unsoundness of this argument resides in the absurdity of the first premise. What we have here is the conjunction of two a posteriori premises: the unity of all knowledge, and reducibility to science as a necessary and sufficient

condition for the real existence of entities. I cannot see what could possibly count as evidence in support of either claim. It is, in any case, strongly counterintuitive to suggest that the Beethoven Violin Concerto does not exist.[31]

It thus remains to be seen how the claim that there are no beliefs can be validly deduced from the claim that folk psychology is false. In a recent essay, Stephen Stich has argued that the claim that there are no beliefs *cannot be* demonstrated to be true. Stich suggests that the inference is only valid relative to some theory of reference. On a very simple version of the descriptive theory of reference, evidence that folk psychology makes false claims about belief would be sufficient to prove that what is referred to as 'belief' in everyday use does not exist. The argument for the non-existence of belief thus takes the form:

i. Either common-sense assumptions about beliefs are true, or 'beliefs' as the term is employed by common-sense do not exist;

ii. At least some of the common-sense assumptions about beliefs are false; therefore

iii. 'Beliefs' does not exist.[32]

The first premise here is the simple descriptive account of reference. There is an ambiguity regarding how much of common-sense would need to be wrong in order for 'belief' to fail to denote. I presume that on a very simple account, *any* falsity in the

claims being made by common-sense would be sufficient for failure to denote. Whether the second premise is true or not is an empirical matter; for the Churchlands, it is regarded as being true, so that the descriptive account of reference serves to demonstrate the truth of the conclusion. However, as Stich points out, this is a theory of reference which will lead to catastrophic elimination: the history of science suggests that all current theories are likely to be false to some degree. By implication, the entities over which all scientific theories currently quantify do not exist.

One popular alternative to the descriptive theory is Kripke's 'Causal-Historical' theory of reference. As Stich points out, however, while the descriptive theory renders the eliminativist claim that there are no beliefs trivially true, the Causal Historical theory makes the claim trivially false. This arises due to the fact that, following its original dubbing, the theory allows for adjustments to the extension of a term - these adjustments being nonetheless reference-preserving. What this entails is that no evidence of the falsity of folk psychology could serve to prove that folk psychological 'beliefs' do not exist; the extension of the term can be amended in the light of any such empirical evidence. This would appear to be sanctioned by Patricia Churchland's co-evolution strategy - I take it that both the central theoretical claims, and the extensions of the proprietary terms of the sciences might 'evolve'. If so, then Patricia Churchland's position entails a much more liberal theory of reference than the catastrophic descriptive theory. The triviality of the falsity of folk psychology arises from the fact that it would appear that no empirical evidence could ever serve to entail the non-existence of beliefs: the defender of beliefs can amend the extension of the term, while maintaining that the amended and new terms have identical reference.

The claim that the alleged falsity of folk psychology entails that there are no beliefs must thus be indexed to some theory of reference which will avoid the dangers of either pan-eliminativism or the frustration of the eliminativist project. In the absence of any such proposed theory from the eliminativist, I propose to consider eliminativism as a thesis concerning explanation, rather than as a straightforwardly ontological thesis. [33]

[1] Paul M. Churchland, *Matter and Consciousness - A Contemporary Introduction to the Philosophy of Mind*, Revised edition, MIT Press, Cambridge Mass., 1988, p.45

[2] I propose to accept, for present purposes, Churchland's term 'folk psychology' for explanation which makes use of belief- and desire-ascriptions. In the next chapter, I will consider the question of the status of 'folk psychology'.

[3] If an event is the instantiation of properties at a time, then type reduction is sufficient for token reduction - but it is not necessary, so that token reduction does not entail type reduction.

[4] Kathleen Lennon and David Charles, 'Introduction', in David Charles and Kathleen Lennon (eds.), *Reduction, Explanation, and Realism*, Clarendon Press, Oxford, 1992, p.2.

[5] Lennon and Charles, *Reduction, Explanation, and Realism*, p.5.

[6] Despite their open contempt for metaphysics, eliminativists characteristically make a number of assumptions which are metaphysical: that the universe is layered, with entities and properties occupying various levels in a hierarchy of explanation; that physics is fundamental - and hence comes at the bottom (hence the demand that mentalistic explanation be vindicated at a more basic level); causal closure at the physical level; etc.

[7] Lennon and Charles, *Reduction, Explanation, and Realism*, p.3.

[8] If we assume that the five conditions for reduction are jointly necessary and sufficient, epiphenomenalism will thus entail the impossibility of reduction.

[9] Charles and Lennon cite a further motivation for reduction: '... the view that real causation takes place at the physical level.' Once again, this is an a posteriori and contingent claim - and cannot be used in an argument for reduction to the physical, on pain of begging the question against the mental.

[10] PS Churchland & TJ Sejnowski, 'Neural Representation and Neural Computation' in William G. Lycan (ed.), *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Mass., 1990, p.229.

[11] Patricia Smith Churchland, *Neurophilosophy*, MIT Press, Cambridge, Mass., 1986, p.373.

[12] Paul M. Churchland, 'The Continuity of Philosophy and the Sciences', *Mind and* Language, Vol. 1, No. 1, 1986, p.6.

[13] quoted in Tom Sorell, *Scientism - Philosophy and the Infatuation with Science*, Routledge, London, 1991, p.6.

[14] I will later argue that eliminativists sometimes fail to distinguish between these two theses - so that a catastrophic form of eliminativism emerges, as all entities and events and processes which cannot be reduced become subject to elimination.

[15] Sorell, *Scientism*, p.4.

[16] Churchland's post-Quinean epistemological convictions - which play a key role in rendering eliminative materialism possible - will constrain his claiming that any such a priori argument *is* available. In its absence, all that is available is induction from past scientific success. But this doesn't support the claim that science is unified - so that speculative optimism regarding the future unifying trend of scientific explanation is the basis for much of what follows.

[17] Stephen P. Stich, *From Folk Psychology to Cognitive Science - The Case*

Against Belief, MIT Press, Cambridge, Mass., 1983, p.222. Stich later resiles from this view, as I will show later in this chapter.

[18] Paul M. Churchland,'Postscript: Evaluating Our Self-Conception', appended to his 'Eliminative Materialism and the Propositional Attitudes' in Paul K. Moser and J. D. Trout (eds.), *Contemporary Materialism - A Reader*, Routledge, London, 1995, p.172. Here and elsewhere Churchland appears to assume that basic physics is free of ceteris paribus clauses (or that, at least, the prevalence of hedging diminishes as one descends closer towards the basic physical level). This implicit assumption is left undefended. Interestingly, however, a standard proposition from physics which is widely utilised in elementary philosophical texts: 'water boils at 100° celsius', requires hedging, as water will boil at higher or lower temperatures at non-standard atmospheric pressure.

[19] Paul M. Churchland, 'Postscript', Moser & Trout (eds.), p.178.

[20] John R. Searle, *The Rediscovery of Mind*, MIT Press, Cambridge, Mass., 1992, p.113.

[21] Paul M. Churchland, 'On the Nature of Theories: A Neurocomputational Perspective', in his *A Neurocomputational Perspective - The Nature of Mind and the Structure of Science*,
MIT Press, Cambridge, Mass., 1989, p.153.

[22] op. cit., p.154.

[23] Presumably Churchland would not express his position in this way: I take it that the relationship of entailment makes necessary use of some form of quasi sentence-like entities. This is an example of how eliminativism can extend well beyond the elimination of simple folk psychological explanation; philosophical logic may be endangered as a consequence of the elimination of sentences.

[24] This empirical study of cognitive processes is 'naturalised epistemology'. Here, as elsewhere, Churchland bases his approach on Quinean assumptions.

[25] My contention is that Churchland does not have the notion of 'theory' that he needs: his PDP account (see ch.4) *embodies* theories as patterns of synaptic weights in the brain. While this allows for a PDP account of the operation of competing theories (identical inputs to the brain being subject to differing patterns of processing activity), it cannot accommodate intertheoretic reduction (see footnote [21], this chapter). Churchland's discussion of reduction comes in his early (pre-PDP) work, where he can avail himself of the classical notion of a theory as being a set of sentences. On this account, even if one rejects the metaphysics which underlies eliminativism, one can at least see how reduction arises. But sets of synaptic weights cannot accommodate the logical and semantic relationships which Charles and Lennon identify. This observation anticipates my central claim in the latter part of the thesis: accommodating logic, epistemology, ethics and the phenomenological will result in eliminative materialism being hopelessly modest; adjusting eliminative materialism to render it more radical in its substantive claims will result in its losing its capacity to accommodate these basic concerns of philosophy. An account of 'theory' which is compatible with radical eliminative materialism, but which cannot accommodate intertheoretic reduction - and, ipso facto, elimination - is self-defeating.

[26] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes, in Lycan (ed.), *Mind and Cognition*, p.208.

[27] Paul M. Churchland, 'On the Nature of Theories', p.153. This condemnation is

admittedly general, but it does seem to apply to the predictive/ explanatory account as well as the sentential model.

[28] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', p.206. Churchland here drops the earlier caution which leads him merely to speculate that such an outcome is likely.

[29] I recognise the difficulty in stating clearly what it is for something to exist, and that the status of the existence of a concerto will differ in potentially problematical ways from that of a violin. Nonetheless, when a musicologist claims that a tenth Beethoven symphony exists, it seems wildly counter-intuitive to deny this *on physicalist grounds*.

[30] Stephen Stich, *From Folk Psychology to Cognitive* Science, p.222.

[31] This raises general problems in formulating the physicalist thesis. If we define physicalism as the thesis that for all *(x)*, if *(x)* exists, then *(x)* is physical, this entails that if *(x)* is not physical, then *(x)* does not exist - so that it may well be that concertos, together with numbers, and propositions, do not exist. The alternative would appear to be either to enter a number of caveats which will redeem both physicalism and concertos, or to define physicalism as the thesis that a complete account of what exists, and the nature of what exists, could in principle be offered by the physicist. Here again problems arise, however: what is the principle, being alluded to? It cannot, on pain of circularity, be the ontological claim that for all *(x)*, if *(x)* exists, then *(x)* is physical. Furthermore, this latter definition is a clear case of scientism: why ought we to defer to the future physicist as the arbiter of what exists?

[32] I have put 'belief' in quotation marks to leave open the possibility - which I take it that Churchland would concede - that cognitive science may provide a theory of belief which is distinct from the common-sense conception, and does not suffer from the alleged errors of the latter. On the simple descriptive account of reference, these would - terminology notwithstanding - be two separate entities (i.e. it would not be the case that both common-sense and cognitive science were discussing the same entity, but common-sense was wrong about the entity.

[33] I am thus accepting, for present purposes, Charlton's claim that 'if physicalism is to define itself in opposition to dualism, the idea it opposes is not that the universe contains two different sorts of entity, but that there are two irreducibly different sorts of explanation' (William Charlton, *The Analytic Ambition*, Blackwell, Oxford, 1991, p.127). It should be noted that Charlton's definition of physicalism is in fact a definition of physicalism, *together with the unity of science thesis*. Fodor insists upon the autonomy of psychology - and is thus, on Charlton's account, a dualist (a characterisation which Fodor would, I take it, reject)

CHAPTER 2 WHAT IS FOLK PSYCHOLOGY?

In presenting his case, Paul Churchland typically considers the empirical question of whether it is likely that folk psychology will eventually be eliminated by neuroscience. [1] The question of the likelihood of elimination conflates two types of question: broadly a priori questions regarding the theoretical status of folk psychology and the ontological status of the entities over which folk psychology generalisations quantify; and speculative questions regarding the likely future development of science. Churchland's view is that the latter - speculative - category is the proper domain for philosophical work, traditional a priori philosophy being effectively worthless:

> There is no ... accumulated compendium of important a priori truths ... and this despite that fact that philosophers have been talking and theorising about them for over twenty five centuries.[2]

Speculating about future scientific discovery - and hence about the possible elimination of folk psychology - is, by contrast, precisely the kind of task which is proper to philosophy:

> The philosopher is just another theorist, one whose bailiwick often places him or her at the earliest stages of the process by which proto-scientific speculation slowly develops into testable empirical theory ... the academic discipline of philosophy typically focuses on domains so far unconquered by any mature science.[3]

This statement of method does much to explain Churchland's approach to his work, which is consistently inductive, empirical, and scientistic - and dismissive of argument based on conceptual analysis. I nonetheless propose to separate what I take to be the

non-philosophical question of the likely future development of neuroscience, from the philosophical questions of the theoretical status of folk psychology, and the ontological status of those entities over which it quantifies. This separation marks the distinction between the empirical and speculative question of the likelihood of the elimination of folk psychology, and the question of the *eliminability* of folk psychology. Eliminability is a necessary but not sufficient condition for future elimination, so that a refutation of the thesis that folk psychology is eliminable is a fortiori a refutation of the thesis that future elimination is likely. Churchland opens his essay 'Eliminative Materialism and the Propositional Attitudes' with the elimination claim:

> Eliminative materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience.[4]

Mere eliminability does not entail this outcome, being a condition which obtains in principle; empirical considerations may militate against elimination in practice. Elimination in practice would require the satisfaction of other conditions, such as a change in the procedures for socialising children into linguistic practice, and the development of a scientific theory which fulfilled all of the tasks currently performed by folk psychology. An outcome whereby folk psychology was eliminable - but retained on the purely instrumentalistic grounds that these practical conditions could not be met - would be a victory for eliminative materialism (albeit of a more limited

nature than that aspired to by Churchland). I thus propose to concentrate on the philosophical question of eliminability in principle.

Churchland's central claim, as established in my last chapter, is that its putative failure to reduce entails the eliminability of folk psychology. There are two suppressed premises at work in this argument:

> that folk psychology is a theory; and
>
> that there is no basic mental level.

The first of these conditions for eliminability - which has been referred to as the 'theory-theory' - is a condition the truth of which is necessary for eliminability, as, by definition, if folk psychology is not a theory then it cannot be a candidate for intertheoretic reduction. In chapter one, I outlined five conditions for scientific reduction, the first of which is the derivation of higher-level laws. In terms of Churchland's position in his essay 'Eliminative Materialism and the Propositional Attitudes', folk psychology meets this condition. In this essay, Churchland alludes to:

> ... the expression of generalisations concerning the lawlike relations that hold among propositional attitudes. Such laws involve quantification over propositions, and they exploit various relations holding in that domain.[5]

If folk psychology is not a theory, then the first condition for scientific reduction is not met: if it is sufficient for some conceptual framework being a theory that it generates putative laws, then it is necessary for laws being generated that the conceptual

framework be a theory. It is unclear how the later, post-logical empiricist Churchland could with consistency support the 'derivation of higher-level laws' condition for reduction. Laws are intrinsically capable of sentential presentation: the derivation of such entities from theories - where theories are not themselves intrinsically capable of sentential presentation appears problematic. The eliminative materialist case thus requires that folk psychology be a theory which fails to reduce to some putatively more basic physical level. In the remainder of this chapter, I propose to consider the case for construing folk psychology as a theory[6].

In his essay 'Eliminative Materialism and the Propositional Attitudes', Churchland presents folk psychology as a set of generalisations, examples of which are:

(i) $(x)(p)$ [(x fears that p) then (x desires that not-p)]

(ii) $(x)(p)$ [(x hopes that p) & (x discovers that p) then (x is pleased that p)]

(iii) $(x)(p)(q)$ [((x believes that p) & (x believes that (if p then q))) then (barring confusion, distraction, etc., x believes that q)]

(iv) $(x)(p)(q)$[((x desires that p) & (x believes that (if q then p)) & (x is able to bring it about that q)) then (barring conflicting desires or preferred strategies, x brings it about that q)].[7]

Here we have states which are related: to each other (e.g. fear causing desire); to external inputs (e.g. the state of affairs which constitutes the content of (p)); and to actions (e.g. x bringing it about that q).

Given his need to characterise folk psychology as a theory, Churchland is providing a question-begging account here: the formal presentation of what appear to be putative laws, with the inclusion in (iii) and (iv) of ceteris paribus clauses, and the analogy which Churchland draws with laws such as:

> $(x)(f)(m)$ [(x has a mass of m) & (x suffers a net force of f) then x accelerates at f/m.[8]

appear calculated to persuade the reader that the case for the theoretical status of folk psychology is both simple and compelling:

> Not only is folk psychology a theory, it is so *obviously* a theory that it must be held a major mystery why it has taken until the last half of the twentieth century for philosophers to realise it.[9]

In his more recent work, Churchland would, as earlier noted, eschew the presentation of a theory as a set of sentences. This sentential account would entail that those who are adept at using the theory would perform cognitive processes on the sentences (or on some physical correlates of the sentences), deducing the explananda in (i) - (iv) above as a result. As already noted, such an account is at odds with Churchland's connectionist account of cognition, and hence with his account of what a 'theory' is. This has the interesting consequence that, on Churchland's account, folk psychology is so inept as to misrepresent itself, by falsely portraying cognition as the manipulation of propositional attitudes.

In a recent postscript to 'Eliminative Materialism and the Propositional Attitudes', Churchland alludes to 'the epistemology that makes [eliminative materialism] possible. [10] Churchland doesn't make explicit the connection between this 'contemporary epistemological perspective' and the possibility of eliminativism, but I take it that the intended connection is that, if true, the epistemological claims will entail the truth of the claim that folk psychology is a theory. That both Churchlands press this latter claim with such insistence is due to the fact that any sound argument to the conclusion that folk psychology is not a theory would in turn be a sound argument for the falsity of the central claim of eliminativism. Patricia Churchland *defines* eliminativism in terms of three claims - the first of which is 'that folk psychology is a theory'.[11] Notwithstanding the Churchlands' denial of the value of a priori philosophy, any proof that folk psychology is not a theory would, on Churchland's own account, entail the falsity of eliminativism without the need for any empirical (or 'neurophilosophical') research.

The epistemology which Paul Churchland cites as performing the task of making eliminativism possible is expressed succinctly. It is 'the realisation that all of human knowledge is speculative and provisional ...'

This view of what knowledge consists in appears impossible to reconcile with Churchland's vigorous attack on the logical empiricist view of a theory as a set of sentences This attack constitutes part of Churchland's attack on folk psychology, with its '...crudely linguaform conception of theories as sets of sentences'.[12] If knowledge is to be 'speculative and provisional', however, this would appear to necessitate that

there be a medium for knowledge - and that this medium be at least capable of *being characterised* in sentential form. It would appear prima facie to be a necessary condition for a set *(x)* to have the property of being speculative that *(x)* be at least expressible as a set of sentences. If this is the case, then there can be no radical error in so characterising the set, irrespective of however else it may be characterised. Churchland thus appears to have a dilemma: the epistemology which, on his own account, makes his metaphysics possible, is inconsistent with one of the central assumptions of the metaphysics. The claimed irreducibility of folk psychology rests in part on its allegedly false characterisation of what possession of a theory consists in - yet that same false assumption is at work in the claim that folk psychology is a theory, and thus eliminable.[13]

Jerry Fodor sets out an argument from Quinean holism to the theory-ladenness of perception, and cites Churchland as an example of a philosopher who has drawn this conclusion from Quine's argument in 'Two Dogmas of Empiricism'.[14] Fodor's explanation of Quinean holism is again inconsistent with Churchland's recent view of what it is to be a theory, however. The account which Fodor offers identifies the semantic and logical relations which must obtain in order for the holist claim to be true: '... inferential relations, evidence relations, and so forth.'.[15] Once again, the presence of such relations necessitates a medium which can at least be characterised in sentential form. Quine's holistic conclusion, regarding our beliefs, that 'revision can strike anywhere' is, it would appear, the basis for the 'contemporary epistemological perspective' in Churchland's paper. If *all* knowledge is speculative and provisional, then - if a theory is defined as a set of claims which is speculative and provisional -

then folk psychology is a theory, and Churchland's immediate claim is secure. Churchland cannot, on his recent account of 'theory', avail himself of this Quinean route to the theory-theory, however, on pain of inconsistency.

The fact of Churchland's argument for the theoretical status of folk psychology being inconsistent with his other claims does not of itself secure the falsity of the claim that folk psychology is a theory. There is, however, an objection to the claim that folk psychology is speculative and provisional, which could be pressed even if Churchland were to relinquish his neurophilosophical account of theories. Churchland makes passing reference to the possible counter-claim that our capacity for introspection yields incorrigible knowledge. No argument is advanced against this Cartesian claim, other than that '...it may seem Palaeolithic and regrettable to some of us ...'.[16]

Two points need to be made about this cursory dismissal of what is an extremely telling argument against the theory-theory. Firstly, the suggestion that the longevity of a claim counts against its truth is an eccentric claim. Secondly, and more importantly, when one considers what Churchland is here denying, the counterintuitive nature of the denial is very considerable. That, as I write this, I believe that I am presenting a critique of Churchland's position, and hope that this critique be effective, and that my desire to make the critique effective causes me to continue writing, is at once an instance of the idioms of folk psychology being utilised in explanation, and incorrigible. It is interesting that Churchland entitles this paper 'Evaluating our Self Conception'. Contrary to what this suggests, his writing seems consistently to adopt a third-person perspective. I fail to see how such a perspective, however well informed

by neuroscience, could plausibly correct my own appraisal of my current mental state. I conclude that, even if Churchland could with consistency argue for the theoretical status of folk psychology on the grounds of the speculative and provisional nature of all knowledge, the claim, when applied to self knowledge gained via introspection of one's current mental states, is in any case false - so that the argument for the theory theory will require some alternative basis.

Churchland has suggested that the controversy surrounding the putatively theoretical status of folk psychology is 'best addressed by rehearsing the history of this notion'. [17] In place of Quine, Churchland here identifies Wilfrid Sellars as having presented the first explicit portrayal of our self-conception as being akin to a theory - in his paper 'Empiricism and the Philosophy of Mind'. Sellars' paper is interesting in that the relevant passages for Churchland's purposes are based on a story which Sellars invents to describe the imaginary development of prehistoric folk psychology. In place of speculation about the future, Churchland is here basing his claims on speculation about the past.[18] Having summarised Sellars' story, Churchland, under the subheading 'Development of the Idea', alludes to 'the *fact* that our conception of the semantic properties of thoughts is derivative upon an antecedent conception of the semantic properties of overt declarative utterances' (my emphasis).[19] Notwithstanding Churchland's nihilistic view of truth, we cannot allow that Sellars has established that this is a 'fact' - especially given that this alleged 'fact' is the basis for the conclusion that our self-conception is theory-like, a claim the truth of which is in turn a constraint upon the truth of the central thesis of eliminativism.

In his paper, Sellars considers the 'classical tradition', according to which

> ... there is a family of episodes, neither overt verbal behaviour nor verbal imagery, which are *thoughts*, and that both overt verbal behaviour and verbal imagery owe their meaningfulness to the fact that they stand to these *thoughts* in the unique relation of 'expressing' them. These episodes are introspectable.[20]

Sellars' criticism of this classical view is that it assumes that such thoughts 'could not occur without being known to occur'.[21] Sellars' position is that, as it is empirically false that there is such self-intimation of thoughts, the classical view is false. It is unclear how the claim that thought is not self-intimating entails what I take to be the desired conclusion - viz. that our knowledge of our internal cognitive states is thus provisional, and hence, in the sense required, 'theoretical'. Sellars' proposed emendation to the classical view allows for 'privileged, but by no means either invariable or infallible access'.[22] Two points must be made here: the Freudian-inspired observation that not all beliefs and desires are self-intimating does not entail that no thoughts are self-intimating; and secondly, the infallibility to which Sellars here alludes appears to be infallibility with regard to the presence of thoughts, rather than with regard to their content, or the attitude which one bears towards that content. His observation, even if true, would thus appear to be consistent with much - or most - thought being both self-intimating and possessed of first-person incorrigibility. I can be certain of the content of my occurrent and conscious propositional attitudes. Sellars goes on to present his case on the apparent assumption that no thought is self-intimating, with his question: 'in what sense can these episodes be "inner" if they are not immediate experiences?'.[23] It is at this point that he

presents 'a myth of my own' - the science fiction speculation to which Churchland alludes .

Sellars' story describes a primitive society of 'our Rylean ancestors'. They have developed a language, but have as yet no conception of the inner states which Sellars has set himself the task of investigating. As a consequence of this deficiency, Churchland suggests:

> They can explain some human behaviours, but only very few. Being linked to a set of operationally defined dispositional concepts, they have no conception of the complex dance of occurrent internal states driving human behaviour, no conception of the internal economy that is just waiting to be characterised by a full-blown theory of human nature.[24]

Sellars' story now involves an individual - Jones - who develops a theory of the internal states. Sellars prefaces this with an account of what it is to construct a theory:

> ... to construct a theory is ... to postulate a domain of entities which behave in certain ways set down by the fundamental principles of the theory, and to correlate ... complexes of these theoretical entities with certain non-theoretical objects or situations; that is to say, with objects or situations which are either matters of observable fact or, in principle at least, describable in observational terms.[25]

Sellars' preamble has in my view failed to establish the need for *all* of the entities quantified over by folk psychology (i.e. all of the internal states) to be the subject of such theoretical postulation. Even if we are to accept the existence of Freud's subconscious internal states, some does not entail all - and no case has been made for the claim that *no* internal state is given immediately to introspection. There is,

however, a more serious problem with Sellars' account of what Jones is doing in forming his theory of internal causal states. These are held to be 'postulated'. 'Postulation' raises a problem for any Rylean account - and this problem infects Sellars' myth: how can postulation proceed in the absence of just those internal states which are being postulated into existence by Jones - indeed, what can 'postulation' be, if not an internal event of the type which is being brought into question? There is an incoherence in the claim that one can 'postulate' in the absence of such internal cognitive states as belief. The internal states which Sellars is here calling into question are, surely, a *precondition* for the formulation of theories. If the presence of internal cognitive states is a necessary condition for postulation, then the act of postulation is sufficient to demonstrate the existence of internal cognitive states. The claim that we possess such inner states cannot then be itself a theoretical claim. Churchland appears not to have noticed this difficulty, as his summary of Sellars' myth also utilises the term 'postulates' - and adds a further intentional term: 'assigns':

> ... in short, Jones postulates the basic ontology of our current folk psychology, and assigns to its elements their now familiar causal roles, much to the explanatory and predictive advantage of everyone who gains a command of its concepts.[26]

Sellars' story concludes with Jones taking as his model for these postulated, theoretical entities, overt verbal behaviour - hence Churchland's earlier-noted reference to the putative derivation of the semantic properties of thought from the semantic properties of our public language. This account is congenial to Churchland's own position in two respects: it appears to account for the error in the putatively erroneous 'linguaform conception' of folk psychology, and for the origin of that error; and - if correct - it

would dispose of the problem which the intentionality of thought poses for the naturalist. Unfortunately for Churchland's account, the problem earlier mooted - that postulation already presupposes the categories of folk psychology - carries over to the semantics of public languages. Surely for any linguistic utterance to have meaning, it must have meaning *for* or *to* some cognizer - and a necessary condition for this being the case is the prior possession of thoughts which have meaning. In the same way in which Jones could not postulate or assign in the absence of internal cognitive states, nor could he make sense of the language which Sellars and Churchland propose to take as an 'explanatory primitive' which will endow thoughts with meaning.

Finally, the notion that we must somehow *learn* 'to make spontaneous first-person ascriptions ... which are strongly consistent with the ascriptions made on purely explanatory or third-person criteria'[27] again seems to reverse the actual order of explanation. That one must learn a public language in order to have a medium for reporting to others one's self-intimating and internal first-person states is undoubtedly the case - but this does not entail what Churchland here apparently takes it to entail: that third-person ascriptions are explanatorily primary. Once again, the actual order is surely that one first has experience of one's first person states, then learns to express them as one acquires a language, and to ascribe similar states to others.

Under the subheading 'Consequences of the Idea', Churchland takes the Sellars story to entail that 'introspective knowledge is denied any special epistemological status', on the grounds that '[it] is hostage to the quality of the background conceptual scheme in which [it is] ... framed'.[28] As Churchland's own writing is entirely from the third

person perspective, he requires some justification for this relegation of the first-person perspective, and the claim that all perception is theory-laden is, I take it, the justification. As Sellars has not, in my view, demonstrated the theory-ladenness of all perception (and thus of all perception of one's own occurrent mental states), I conclude that Churchland needs some alternative justification for the claim that:

> ... introspective judgements ... turn out not to have any special status or integrity ... and introspective judgement is just an instance of an acquired habit of conceptual response to one's internal states ... contingent on the integrity of the acquired conceptual framework (theory) in which the response is framed.[29]

Churchland takes as a second of the consequences of Sellars' idea, that 'the traditional mind-body problem emerges as a straightforward scientific question - as a question of how the theoretical framework of folk psychology will turn out to be related to whatever neuropsychological theory might emerge to replace it'.[30] Churchland here rather begs the question against folk psychology by alluding to its 'replacement', but for my present purposes what is of greater concern is the fact that Sellars' theory-ladenness thesis is again being taken to entail the conclusion which eliminativism requires. The argument at work here appears to be:

i. all knowledge is based on perception;

ii. all perception is mediated by theoretical - and hence speculative - assumptions; therefore

iii. all knowledge is speculative (and hence 'theoretical')

iv. therefore our self-knowledge is theoretical.

The first premise is a straightforward empiricist assumption. While the fact that his subscription to this view will account for Churchland's scientism, and rejection of philosophy as an a priori discipline, it remains to be proved that (i) is in fact true. Premise (ii) is the Sellarsian assumption which I have already argued is not proven. A further difficulty emerges for Churchland when one asks how the conclusion is to be put to use by the eliminativist. If all knowledge is speculative, then the normative character of knowledge is lost: why should one accept one knowledge claim in preference to another? Churchland's position here seems to be self-defeating: one can only accept his conclusion if one accepts it as being true, in contrast to its provisional and speculative opponents.

Having drawn this conclusion from the Sellars story, Churchland then goes on to suggest that a failure on the part of folk psychology to reduce to neuropsychology would have the result that:

> the successor theory will then displace Jones's antique theory in our social and explanatory practices, and the ontology of folk psychology will go the way of phlogiston ... [31]

That no such outcome would be guaranteed by the truth of the theory-theory, together with the irreducibility of folk psychology , should be clear: even if folk psychology is a theory, and there is no basic mental level, it remains to be proved that irreducibility entails eliminability - the claim which non-reductive physicalists will reject. But the quotation also shows Churchland making an inferential leap from eliminability to actual elimination - which would, as already urged, require the satisfaction of certain practical

conditions such as a change in the practice for socialising children into explaining human action. Finally, Churchland assumes that the form of elimination which would be the outcome is ontological elimination - which, as I suggested earlier, is the strongest form of elimination, and one which is only entailed by elimination of folk psychology explanation if one assumes the truth of the simple descriptive account of reference.

I have now considered two possible routes to Churchland's conclusion that folk psychology is a theory. Churchland cannot avail himself of the Quinean route, as this is inconsistent with his own current position on theories. The Quinean route appears also falsely to characterise all first-person knowledge as speculative.[32] Finally, I have argued that the Sellars route to the theory-theory (the categories of folk psychology are postulates - hence theoretical entities, and so folk psychology is a theory) fails on two grounds. Firstly, the position suffers from pragmatic incoherence, in that it assumes that one can postulate in the absence of internal cognitive states. Secondly, the false claim is apparently made that, as some mental states are not self-intimating, none are - so that postulation is required.

In his essay 'Evaluating Our Self-Conception', Churchland rehearses various criticisms which are levelled against eliminativism, and attempts to rebut each. One such criticism is separately levelled by Wilkes and Hannan, and is based on the observation that:

> ... the conceptual framework of folk psychology is used for a vast range of "non-scientific" purposes beyond the prototypically "theoretical purpose of describing the ultimate nature of human psychological organisation.[33]

If Wilkes' and Hannan's objection stands, then as folk psychology is not then a theory, it is not eliminable.

Churchland accepts the case being made here for the practical function of folk psychology , 'yet the conclusions drawn therefrom betray a narrow and cartoonish conception of *what theories are* and *what they do'* (my emphasis).[34] In what he apparently takes to be a contrast with the 'cartoonish' account of theory of his opponents - 'abstract propositional description, invented for the purpose of deep explanation far from the concerns of practical life' - Churchland presents 'theory' as being in continual use in our everyday lives: thus the jazz musician uses musical theory, the carpenter geometry, and the blacksmith metallurgy, mechanics and simple thermal physics. Despite his recognition that there are two questions to be addressed with regard to theories: 'what they are', and 'what they do', it is, however, once again difficult to see from this passage what Churchland takes a 'theory' to be, if not a framework of abstract propositions for the purpose of explanation.[35] Churchland evades this question by concentrating on the 'what they do' question - citing examples of the practical abilities of the individual who *has* command of a theory, and the relationship which he claims to obtain between the development of skills, and mastery of theory:

> ... what learning a theory amounts to [is] much less the memorising of doctrine and much more the slow acquisition and development of a host of diverse *skills* ...[36]

The development of practical musical skills may be made possible by, and reinforce the musician's grasp of, musical theory. As Churchland himself notes: '... sustaining

enhanced practice is what theories typically do'.[37] When one returns to the 'what they are' question, however, the fact of their sustaining skills does not entail that theories are not most appropriately thought of as 'abstract propositional descriptions invented for the purpose of deep explanation'.[38] Churchland's characterisation of theory as intimately related to practical skill is in any case too narrow. It is difficult to see how much philosophical theory could be accommodated by this account - for example, Plato's Theory of Forms, which appears to be a paradigmatic case of 'abstract propositional description invented for the purpose of deep explanation' .

Churchland thus rejects the claim of Wilkes and Hannan that folk psychology is not a theory, and hence not eliminable. His position is that their characterisation of what folk psychology does - serving practical needs - is consistent with its being a theory. In advancing his pragmatic account of theory acquisition, Churchland cites Kuhn in support of this position: 'our best ... [account] of what learning a theory amounts to' is, he contends, contained in Kuhn's *The Structure of Scientific Revolutions'*.[39]

Kuhn's account of scientific progress as a series of paradigm shifts is clearly congenial to Churchland's own position, which proposes a radical revision of our self-conception. In *Scientific Realism and the Plasticity of Mind*, Churchland locates Kuhn's text along with the work of Quine in the tradition of Naturalised Epistemology - the epistemology, that is, which, on Churchland's own account, makes eliminativism possible. In Kuhn we thus have a third putative source of support for the claim that folk psychology is a theory, and thus for the truth of the first premise in the argument

for eliminability. In *Scientific Realism and the Plasticity of Mind*, Churchland provides scant support for the naturalised epistemology which he attributes to Kuhn:

> One can be impressed with the poverty of current a priori epistemology when confronted with the intricate details and grand dramas of our actual theoretical development (cf. Kuhn).[40]

For some insight into the alleged 'poverty' of conventional a priori epistemology, one may turn to Patricia Churchland, who in her essay 'Epistemology in the Age of Neuroscience' claims to speak for both herself and Paul Churchland.[41] Patricia Churchland here presents 'the fundamental epistemological question from Plato onwards' as being 'how is it possible for us to represent reality?'.[42] To this premise is added the physicalist premise that 'it is ... the nervous system that achieves [representation of reality]; yielding the conclusion that 'the epistemological question can be reformulated thus: *How does the brain work?*.[43] If this was indeed the central question of epistemology, then it would occasion little surprise that a priori epistemology was impoverished. This conclusion is arrived at by a sleight of hand, however: both of the premises are false, so that a priori epistemology is redeemed.

Patricia Churchland's first premise - that the central question of epistemology is 'how it is possible for us to represent reality?' - is carefully framed in order to beg the question in favour of neurophilosophy. If one takes the fundamental question to be concerned with knowledge, as the etymology of the term would dictate, and if one further takes a necessary condition for knowledge to be justification, then the naturalist is confuted. As Paul Churchland points out:

> 'Ought's not being derivable from 'is's, it would seem that normative epistemology cannot be a purely empirical science.[44]

If epistemology is inherently normative, then this entails the falsity of the naturalist project: if all that exists is natural, and can thus be studied by those methods appropriate to the study of nature, then there can be no accommodation of values - no discovery of what ought to be the case . Jonathan Dancy defines epistemology as 'the study of our right to the beliefs we have'.[45] Here there is not only the normative dimension which Churchland concedes not to be accommodated by science, but in addition, the claim that the target of epistemology is the justification of entities, the existence of which the eliminativist is concerned to deny. I conclude that if there is a case for naturalising epistemology, Patricia Churchland has not made it here.[46]

Naturalism provides a putative route to the irreducibility of folk psychology , rather than to the claim that folk psychology is a theory. If we try to elicit Churchland's argument, contra Hannan and Wilkes, from the truth of Kuhn's thesis to the truth of the theory-theory, the argument seems to be as follows:

i. learning a theory consists in developing skills which are 'internalised';

ii. the folk psychology adept has internalised skills;

therefore folk psychology is a theory.

If this is Churchland's argument, then it is surely invalid: one cannot deduce the answer to the question 'what is a theory?' from the answer to the question 'how is a theory learnt?'. This is, I take it, where the reformulation of epistemology's central question in naturalised epistemology comes into play. If the central question is 'how does the brain work?', then this must entail that the answer to the question 'how is a theory learnt' must be in terms of purely physical processes. The theory must thus be postulated not to have the intentional characteristics of being true or false, or having its components related by logical interrelationships.[47] All that remains is that the behaviour of the system - the skills which it undertakes - is subject to change as a consequence of the 'internalisation' of the theory. Even on this account, it seems that the most that can be deduced from the answer to the question 'how is a theory learnt?' is an answer to the question 'what is it that theories *cannot* be?'. Once again, the question of Churchland's positive thesis on the question of what theories are is left mysterious.

Apart from the invalidity of the argument which I have extracted from Churchland's text, the first premise appears to be empirically false. The 'internalisation' of theories is alluded to by Kuhn:

> ... during revolutions scientists see new and different things when looking... in places they have looked before ... looking at a bubble-chamber photograph, the student sees confused and broken lines, the physicist a record of familiar subnuclear events. Only after a number of such transformations of vision does the student become an inhabitant of the scientist's world, seeing what the scientist sees and responding as the scientist does.[48]

Here we have the theory-ladenness of perception thesis, combined with the notion that the acquisition of a theory involves the recipient in a 'change of world view'. The unit

of understanding is on this account a 'paradigm', rather than a proposition, or set of propositions. As a consequence, the individual may not be able to retrieve from the mind the theory as readily as one would with a set of consciously-known propositions. Nonetheless, the theory is 'internal' in the sense that one cannot but see the world in its terms. Churchland reinforces the point by listing some of the skills putatively involved:

> ... skills of perception, categorisation, analogical extension, physical manipulation, evaluation, construction, analysis, argument, computation, anticipation ...[49]

This list includes tasks putatively performed by folk psychology - such as evaluation, argument and anticipation. As with the earlier-noted comparison between laws of mechanics and 'laws' of folk psychology , this is an argument from analogy: the acquisition of theoretical skills facilitates argument; ability to use folk psychology facilitates argument; therefore folk psychology is a theoretical skill. The falsity of this analogy arises from the fact that Churchland has conflated two sets of preconditions for advancing or understanding an argument in, to take Kuhn's example, nuclear physics. In order to do this one must have a command of the relevant nuclear theory. But prior to this is a command of the idioms of folk psychology. Far from being in competition with it, folk psychology is a prerequisite for scientific theorising: having the belief that (p) is a necessary condition for having an understanding of (p), or - in Churchland's terms - being able to apply the skill which is contingent on having (p). If the first condition for the eliminability of folk psychology is that folk psychology be a theory, then eliminativism falls at the first hurdle: far from being a theory, folk psychology provides the context within which theories are known and learnt and

applied. The eliminability of folk psychology entails the eliminability of all theories - an outcome which even the most catastrophic eliminativist cannot countenance.

One final possible route to the theory-theory arises from the reification of belief.[50] If beliefs are entities, and beliefs are invoked in common-sense explanation, then this might be thought to entail that beliefs are theoretical entities, and the framework within which they figure is thus a theory.[51] If one rejects the reification of belief, then the temptation to view beliefs as theoretical entities is forestalled.

I would suggest that to have a belief is to be disposed to assign the truth value *(T)* to some proposition, in the light of some assumed justification.[52] The question of whether one has a non-occurrent belief is then addressed by the use of counterfactuals - as opposed to neurological investigation. Folk psychology is not a theory, as theories are - as Sellars suggests - consciously created sets of claims, constructed with a view to explaining the phenomena under investigation. *Pace* Sellars, folk psychology is not the result of such conscious creation. Folk psychology does satisfy the condition of explaining human action - but folk psychological idioms are prior to folk psychological explanation. By 'prior to', I mean that one must have a belief - be disposed to assign the relevant truth-values - as a necessary condition of postulating the existence of theoretical entities, and hence constructing a theory. As they are prior to the creation of theory, folk psychological idioms thus cannot be taken themselves to denote theoretical entities, capable of elimination in the light of some subsequent, and putatively superior, theory.

I conclude that there is no sound argument for the conclusion that folk psychology is a theory. If - as the Churchlands themselves claim - the truth of this claim is a constraint on the truth of the central claim of eliminative materialism, then I conclude that Eliminative Materialism is false.

[1] Other physicalists consider scientific psychology to have a stronger claim to be the ultimate best theory of human cognition. For my present purposes, nothing hangs on this dispute.

[2] Paul M. Churchland, 'The Continuity of Philosophy and the Sciences', in *Mind and* Language, Vol. 1, No. 1, 1986, p.8.

[3] Paul M. Churchland, 'The Continuity of Philosophy and the Sciences', p.6.

[4] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', in William G. Lycan (ed.): *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Mass., 1990, p.206.

[5] op. cit., p.209.

[6] There is a further necessary condition for the eliminability of folk psychology: the non-existence of a basic mental level. My ch.6 will argue for an ineliminable and irreducible mental level, so that my claim is that even if folk psychology were to be a theory, it would escape elimination in virtue of this domain of the irreducibly mental.

[7] op. cit.

[8] op. cit.

[9] op. cit.

[10] Paul M. Churchland, 'Postscript: Evaluating our Self-Conception', in Paul K. Moser and J. D. Trout (eds.), *Contemporary Materialism - A Reader*, Routledge, London, 1995, p.169. The fact of this being a recent postscript to the 1981 essay 'Eliminative Materialism and the Propositional Attitudes' - written for inclusion with the essay in Moser and Trout's text - is important for what follows, in that Churchland has revised his view of what is a theory since writing the original essay.

[11] Patricia Smith Churchland, *Neurophilosophy*, MIT Press, Cambridge, Mass.,1986, p.396

[12] from Paul M. Churchland's essay 'Folk Psychology' in Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Blackwell, Oxford, 1994, p.313.

[13] My earlier concern to distinguish between the a priori question on the eliminability of folk psychology , and the empirical question of the likelihood of its eventual elimination, comes into play here. Mere eliminability does not require that folk psychology provide a false characterisation of theories: in 'Eliminative Materialism and the Propositional Attitudes', as already noted, Churchland appears to accept the sentential view of theories, as part of his case for eliminability (FP is a theory, therefore it is in principle reducible, and if not then it is in principle eliminable). His subsequent argument that folk psychology mis-characterises theory is the result of his empirical work in neuroscience - and is intended, I take it, to present evidence to the effect that the future elimination of folk psychology is likely. As with the question of whether irreducibility entails falsity, the issue of whether folk psychology is merely eliminable in principle, or likely to be eliminated in practice, appears to be a debate exclusively between physicalists. In the latter case, Churchland's opponents would appear to be instrumentalists.

[14] In his essay 'Observation Reconsidered', in *A Theory of Content and Other Essays*, MIT Press, Cambridge, Mass., 1990, p.231.

[15] op. cit., p234.
[16] Paul M. Churchland, 'Postscript: Evaluating our Self-Conception', in Moser & Trout (eds.), p.169.
[17] Paul M. Churchland, 'Folk Psychology (2)' in Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, p.308.
[18] Churchland's Quinean-inspired epistemology may come into play here. If all knowledge is speculative and provisional, then this is inconsistent with the simple correspondence theory of truth - hence Churchland's breathtaking claim that: 'it is no longer clear that there *is* any unique and unitary relation that virtuous belief systems must bear to the non-linguistic world' ('On the Nature of Theories', in his *A Neurocomputational* Perspective, p.157). One is tempted to point out that, on the ontological reading of the eliminativist thesis discussed in my last chapter, the fact that there are no beliefs must entail that there are no belief systems, virtuous or otherwise. Setting this objection to one side, it still remains unclear how one is supposed to proceed from this nihilistic position regarding truth - but what Sellars candidly describes as 'a piece of ... anthropological science fiction' (in 'Empiricism and the Philosophy of Mind', in his *Science, Perception and Reality*, Routledge, London, 1963, p.178) presumably cannot be ruled out of consideration as a basis for arriving at a conclusion which is 'virtuous'.
[19] Paul M. Churchland, 'Folk Psychology', in Guttenplan (ed.), p.309.
[20] Wilfrid Sellars, 'Empiricism and the Philosophy of Mind', in his *Science, Perception and Reality*, p.177.
[21] op. cit.
[22] op. cit. p.178.
[23] op. cit.
[24] Paul M. Churchland, 'Folk Psychology (2), p.308.
[25] Wilfrid Sellars, 'Empiricism and the Philosophy of Mind', p.181.
[26] Paul M. Churchland, 'Folk Psychology', in Guttenplan (ed.), p.309.
[27] op. cit.
[28] op. cit. p.310.
[29] Paul M. Churchland, 'Eliminativism and the Propositional Attitudes', in Lycan (ed.), p.208.
[30] Paul M. Churchland, 'Folk Psychology', in Guttenplan (ed.), p.310.
[31] op. cit.
[32] Whether or not this is Quine's own view, this would be e necessary condition for the truth of the claim that, qua speculative knowledge, folk psychology is a theory.
[33] Paul M. Churchland, 'Postscript: Evaluating our Self-Conception', in Moser and Trout (eds.), p. 174 (quotations taken from Hannan, 'Don't Stop Believing: The Case Against Eliminative Materialism', in *Mind and Language* Vol. 8, No. 2. (1993).
[34] op. cit.
[35] I have earlier anticipated my discussion in ch.4 of Churchland's positive account of what a theory is (footnote 25, p.24). The claim that a theory is an embodied set of synaptic weights in the brain is a hopelessly impoverished account of what a theory is, however: the account appears to accommodate various performances on the part of the individual whose synaptic weights they are (the issue on which Churchland here concentrates), but the question of how such a set of weights relates to a theory

which is, for example, in a *book*, is left unexamined. What is wrong with construing a theory as 'abstract propositional descriptions invented for the purpose of deep explanation'? If for Churchland it is merely (as one suspects) that Fodor's language of thought thesis is flawed, and that there are, in consequence, no 'sentences in the head', then Churchland is, as suggested on p.24, throwing out the baby with the bath water here, as this rejected construal of what is a theory is necessary for a coherent account of elimination itself to be possible. If, on the other hand, the synaptic sets map on to theories in books in a 1:1 manner, then the claim that 'there are no sentences in the head' must be taken in a crudely literal way for its truth to be secured, and eliminative materialism is itself left absurdly modest in its central claim.

[36] op. cit.

[37] op. cit. p.175.

[38] op. cit. p.174.

[39] op. cit.

[40] Paul M. Churchland, *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge, 1979, p.124.

[41] Patricia Smith Churchland, 'Epistemology in the Age of Neuroscience', in *The Journal of Philosophy*, 1987, p.545.

[42] op. cit., p.546.

[43] op. cit.

[44] Paul M. Churchland, *Scientific Realism and the Plasticity of Mind*, p.124.

[45] In Ted Honderich (ed.), *The Oxford Companion to Philosophy,* Oxford University Press, Oxford, 1995, p.245.

[46] Churchland's second premise - that the mind is the brain - is, of course, the routinely-presented physicalist assumption which is left unsupported, on the apparent assumption of its obvious truth.

[47] Churchland credits Kuhn's text with 'upsett[ing] my own Logical Empiricist assumptions' , in his *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Mass., 1995, p.272.

[48] Thomas S. Kuhn, *The Structure of Scientific Revolutions*, (second edition), University of Chicago Press, Chicago, 1970, p.111.

[49] Paul M. Churchland, 'Postscript: Evaluating our Self-Conception', in Moser and Trout (eds.), p.175.

[50] I have no particular text or philosopher in mind here: the position which I am setting out often seems to be implicit in the arguments advanced.

[51] I do not claim that Paul Churchland subscribes to this position: his account of what is a theory does not appear to be capable of accommodating theoretical entities.

[52] This justification condition not being met by naturalised epistemology, the reification of belief might be thought to avoid this difficulty: putatively natural theoretical entities cannot have the property of being justified - so that they are misconceived within the framework of folk psychology , which incorporates what is, in effect folk epistemology, as common sense (wrongly, on Churchland's view) assumes the need for knowledge to be justified.

CHAPTER 3 WHAT DOES ELIMINATIVE MATERIALISM PROPOSE TO ELIMINATE?

In his essay 'The Varieties of Eliminativism: Sentential, Intentional and Catastrophic' [1], Andy Clark responds to an essay by Barbara Hannan.[2] Clark draws, in the first instance, a useful distinction between what he terms 'Intentional Eliminativism', and 'Sentential Eliminativism' .Intentional Eliminativism: ' ... questions the correctness of all *descriptions* of our mental states which invoke propositional contents' .Sentential Eliminativism, on the other hand ' ... *need not* balk at such descriptions, and insists only that any actual internal representations in the agent concerned are unlikely to exhibit a comparable language-like structure' (italics in original). Clark's thesis is that eliminativism is confronted with a problem here - for while Sentential Eliminativism is relatively easy to defend, it is arguably not a genuinely eliminative thesis in any significant sense. Intentional Eliminativism *is* genuinely eliminative - but is, perhaps, impossible to defend, and liable to collapse into a much more radical form of eliminativism - 'Catastrophic Eliminativism' - which is incoherent.

Hannan's paper identifies as a weakness of eliminativism the apparent fact of our being rational agents. Given that a rational agent is rational in virtue of adopting beliefs on the basis of appropriate reasons, it cannot, Hannan claims, be wrong to *characterise* the state of such an individual in terms which make use of propositions.[3] To demonstrate the difficulty which any anti-propositional account will face, Hannan cites an example attributed to Pylyshyn:

> Suppose a person ... has come running out of a building. If we agree to see [this person] as a cognizer, and to view her running out of the burning building as a rational act, then it must be admitted that her action might have resulted from a variety of different stimuli. She might have smelled smoke. She might have heard an alarm. She might have seen flames. She might have heard someone yell, 'Fire!' ... This is the feature of rational acts that is sometimes called stimulus independence. A rational act is a response to the meaning or import of the stimulus, not to the stimulus itself; a rational act is not a mere tropism, nor is it a mere conditioned response ... [4]

Hannan then presents 'the crunch': '... unless we see [the person's] causally-relevant states as *representational* ... how are we going to account for the rationality and the stimulus-independence of her act?' (emphasis in original).[5] Hannan's position here seems impregnable to doubt: that, at the very least, the internal state of the individual fleeing the fire must be *open to description* by means of propositions, and the relations of entailment which obtain between these propositions. Despite her apparent endorsement, earlier in the essay, of some form of Quinean epistemology (what Churchland has referred to as 'the epistemology that makes [eliminative materialism] possible'[6]), Hannan surely understates the case, in her claim that: ' ... it is at least arguable that where rational capacities are the explananda, it is necessary that there be propositional attitudes in the explanans'.[7] While it may be arguable what further ontological claims ought to accompany a commitment to the truth of this claim, the claim itself is surely true in virtue of being analytic.

Lynn Rudder Baker regards Hannan's use of analyticity in her argument as both unacceptable to the eliminativist, but also ultimately unsuccessful on its own terms, in opposing the eliminativist claim that 'the best science of mind will not underwrite propositional attitude concepts'.[8] Baker's argument against Hannan's position on

this point is based on her apparent confusion of two issues which Clark in his essay has insisted must be kept separate: the issue of whether there are internal mental states which have semantic properties, and which may thereby be referred to as 'propositional states'; and the issue of whether it is appropriate to *describe* an agent's internal mental states by recourse to propositional attitude ascriptions.[9] Thus, for example, Hannan notes that:

> ... unless cognitive science is cognitive in name only, it had better acknowledge that the internal states it studies possess *some kind* of propositional content ...[10]

This conclusion seems too strong - or at any rate stronger than the claim which is required by her argument. To claim that the states *possess* 'some kind of propositional content' is, it may be argued, to claim that something akin to a language of thought exists. Hannan generates some confusion when she earlier states that:

> 'if the meanings of stimuli [of the individual fleeing the fire] are internally represented, then there are internal states *appropriately characterised as* having propositional content. That is, *on one reading of what a propositional attitude is, there are propositional attitudes.* (my emphasis)[11]

The first sentence of this quotation is, I suggest, unproblematically true. The second sentence, however, seems ambiguous. That the internal states are appropriately characterised propositionally only entails that propositional attitudes exist on a very limited reading of the claim that 'propositional attitudes exist' .In her paper, Hannan shifts between the straightforward claim that propositional characterisation is

appropriate, and the more contentious (and ambiguous) claim that propositional attitudes *exist*.[12]

Baker's criticism of this part of Hannan's paper is based on the inconsistency between Hannan's paper's 'pragmatic billing' and Hannan's claims:

a. that a science of mind will study cognition; and

b. that it is analytic that the study of cognition must yield an account of rationality, which must in turn recognise the appropriateness of propositional description - so that propositional attitude concepts cannot not be 'underwritten' by any science of mind.

The claimed inconsistency resides in the attempts in both (a) and (b) to stipulate a priori the explanans and ontology of science of mind, given Hannan's endorsement of post-Quinean epistemology - which is concerned to deny the possibility of such a priori stipulation.

While Hannan may be open to criticism for her inconsistency, this does not of itself render claims (a) and (b) above false. Baker's more significant criticism is that Hannan's argument rests upon an equivocation on the term 'rational':

> 'If we take 'rational' in a pre-theoretical sense, so that most everyone would agree on which were rational capacities, then [the claim that cognition involves rational capacities] ... is unproblematically true. But in the sense of 'rational capacities' that would make [this claim] ... true, [the claim that rational capacities cannot be

> explained other than by reference to internal states that possess propositional content]... would not be available as a premise in the argument.[13]

Baker's objection is that there is no empirical evidence for this latter claim (i.e. the claim that we cannot dispense with reference to internal states which possess propositional content). As earlier suggested, the difficulty here lies with Hannan's presentation of her case, which may be taken as committing her to the existence of a language of thought, by virtue of her reference to the 'possession' by mental states of propositional content. Had she eschewed use of the term 'possession', then Baker's objection would surely lack force. An alternative presentation of Hannan's argument might be as follows:

i. any science of mind must account for the stimulus-independence of human action;

ii. this account must present action as being rational;

iii. any such account will advert to processes *which may be described* by means of sentences in natural language - so that propositional characterisation of one's mental states (at least) cannot be mistaken.

Premise (iii) above makes no commitment to the actual existence of token propositional or sentential items in the head. In my earlier discussion of theories I suggested that any replacement which Churchland proposes for the sentential model of theory, designed to accommodate his claim that the role of philosophy - and by extension philosophical theory - is speculative, must respect the necessary condition for a set *(x)* to have the property of being speculative that *(x) be at least expressible* as

a set of sentences.[14] Here the same point can be brought to bear in defence of Hannan's position: while it may be thought plausible to argue that neither theories nor human heads actually contain sentences, it is nonetheless a priori that both speculation and accounts of human agency must be open to sentential *description*. Both Hannan and Baker have conceded too much to the eliminativist in accepting the pragmatist epistemology 'that makes [eliminativism] possible' .Thus, for example, Hannan claims that:

> 'we don't know which of our current 'indubitable truths' will turn out to be doubtful after all, in the light of future empirical discoveries and conceptual changes ...'.[15]

As Hannan goes on to point out, this claim is in turn used by the eliminativist in support of the further claim that folk psychology consists of a speculative body of claims - and that, as such, it is a candidate for reduction or elimination (i.e. it underpins the 'folk-theory theory, the truth of which, as already noted, is a constraint on the truth of the central claim of eliminative materialism).[16] Reluctance to endorse the seemingly unfashionable analytic/ synthetic distinction has led in Hannan's paper to a failure to press her - very telling - criticism of eliminativism with sufficient force.

Consistent with her apparent acceptance of the possibility of 'future empirical discoveries and conceptual changes', Hannan anticipates the eliminativists' response:

> '... the eliminativist will protest at this point that I am begging the question; *cognition* and *rationality* are concepts that (like all concepts) could change over time or become outmoded.'.[17]

Rather than press the objection that this is in fact an impossible outcome - it is a priori that cognitive states are rational, and that instances of rational choice are at least open to propositional description, Hannan adheres to the pragmatic line on this occasion:

> '... in the absence of plausible replacements for these concepts ... don't we have ample reason to bet against the eliminativist?'.[18]

This reiterates an earlier suggestion that: 'in the absence of contenders for replacement concepts, it is reasonable to say to the eliminativist ... "Put up or shut up"'.[19]

Two observations can be made regarding this line of attack. Firstly, it is a weak response to eliminativism - presumably necessitated by Hannan's desire to maintain allegiance to post-logical positivist epistemology, despite her de facto use of conceptual analysis to present a potentially very effective problem for eliminativism. Secondly, the argument demonstrates how, by shifting the focus of debate on to the empirical ground favoured by Churchland, and thus denying herself the opportunity to maximise the effectiveness of her critique, Hannan allows her opponent the opportunity of counterattack. Absence of evidence is not evidence of absence, as far as empirical studies are concerned: as Hannan herself notes, the eliminativist will accuse her of begging the question.

In his own response to Hannan's paper, Paul Churchland characterises Hannan's 'put up or shut up' argument as: 'if [folk psychology] is currently the only boat afloat, isn't this ample reason to expect that it will continue to be the only boat afloat?'.[20] Thus presented, the argument is clearly invalid - though for reasons stated earlier,

Churchland will presumably not present his counter-argument in these terms.[21] Churchland writes in entirely figurative terms in this part of his essay, so that his meaning is obscured. Thus, for example, he here promotes a relatively moderate form of eliminativism, which merely 'urges the poverty of our current home', together with 'the pressing need to explore the construction of one or more new ones, and the probability that we will eventually move into one of them'.[22] It is left unclear what it is which the possible replacements would be replacements *for*. If Churchland is here envisaging replacements for the concepts *cognition* and *rationality*, then it remains to be seen what he takes these concepts to *be*. Churchland's earlier-noted allusion to folk psychology in his response to Hannan suggests that what he envisages replacing is the putative theoretical posits of folk psychology. If we take Hannan *not* as arguing for some form of language of thought, but rather for internal states which are *appropriately characterizable* as having propositional content, then Churchland must be taken to be arguing here for the possible future discovery of internal states which are *not* so characterizable. Speculation and the eschewal of traditional epistemology may lead Churchland to regard this is a possibility. But my inclination is here to press the question: in virtue of what would these states then participate in processes which could aptly be described as cognitive or rational states?

I proposed earlier a reconstruction of Hannan's 'rationality' argument which is, I suggest, immune to the criticism levelled against it by Baker:

i. any science of mind must account for the stimulus-independence of human action;

ii. this account must present action as being rational;

iii. any such account will advert to processes *which may be described* by means of sentences in natural language - so that propositional characterisation of one's mental states (at least) cannot be mistaken.

Regarding (i), Churchland's options are either to deny stimulus-independence, or to deny that it is the task of a science of mind to account for stimulus-independence - or to accept the premise. Interestingly, Churchland chooses not to address the 'burning building' example - an example which, I have suggested, presents an especially serious challenge to eliminativism. The denial of stimulus-independence would appear not to be an option for Churchland, who notes, as a 'major flaw' of behaviourism that it 'evidently ignored, or even denied, the "inner" aspect of our mental states'.[23] As a science of mind must surely therefore give an account of the inner causal states which give rise to the woman's fleeing the burning building, I assume that Churchland will endorse the first premise in my reconstruction of Hannan's argument.

Whether Churchland will also endorse the second of my premises - that action must be construed as being rational - is less clear. Churchland opens his essay by characterising the conventional self-conception as 'a rational economy of propositional attitudes'[24] - but appears to envisage only the elimination of the propositional attitudes - construed realistically, as being somehow tokened in the brain. Rationality would presumably thus survive - though rational processes would not be held to consist in manipulations of internally-coded propositional attitudes. Patricia Churchland appears to vindicate this reading[25], when she alludes to our currently 'inchoate understanding of

rationality, which is all we have, until neuroscience and psychology yield a more complete theory of mind-brain function'.[26] I take it that we can thus assume that the future development of neuroscience is not thought likely to result in the elimination of rationality - though, in the spirit of scientific pragmatism, we ought to avoid attempting to stipulate a priori what the explananda of science are in advance of the future development of that science: scientists who embarked on a research programme to investigate alchemy eventually eliminated alchemy as a legitimate part of our conceptual scheme, and such historical precedents are salutary for the Churchlands.

If we assume that Churchland is thus liable to accept - perhaps with some modification - the first two premises, this leaves the third: the acceptability of the propositional *characterisation* of one's inner states. It is here that Clark's 'Sentential/ Intentional Eliminativism' dichotomy comes into play, and Clark's contention that 'eliminativists like Paul Churchland have ... tended to shift around a little between these two positions'.[27] In his rejection of Hannan's 'no existing alternatives' critique, Churchland alludes to 'some very specific and highly promising "replacement concepts" under active exploration'.[28] Churchland proceeds to outline the current research findings in connectionism. Referring to the representational apparatus postulated by this research, Churchland uses scare quotes in referring to 'the "semantics" of the representation' .But the following sentence, with its allusion to 'information-preserving transformation from external world to internal representation' [29] suggests that the eliminativism at work here is (mere) sentential eliminativism: what is being denied is that these putative representations have a sentential structure. Where there is 'information preserving transformation', the subject of this

transformation must be open to description in terms of propositions expressed in natural language. As Clark points out, Churchland reinforces this reading by elsewhere providing just such a description of these intrinsically non-sentential representations:

> 'Is the mouse eating sesame seeds? Or hickory nuts? Is it avoiding a cat? Or a hawk?[30]

It seems, then, that there is no significant difference between Hannan's paper and Churchland's response in terms of what both would be prepared to eliminate: on this occasion, at least, Churchland is a (merely) sentential eliminativist: he denies that, corresponding to the propositional descriptions of inner states, there are sentential (or quasi-sentential) tokens, somehow encoded in the brain. Hannan also denies this - though her unsupported claim that 'propositional attitudes exist', especially if taken out of context, may be thought to suggest otherwise.

Representation in the brain, claims Churchland, involves 'vectorial representations' - intrinsically non-sentential patterns of brain activity. Churchland concludes:

> 'What is important for the issues of this paper is that the relevant sciences have indeed articulated fertile and systematic theories concerning representation and computation in the brain. From the perspective of those theories, the most general and fundamental form of representation in the brain has nothing discernible to do with propositions, and the most general and fundamental form of computation in the brain has nothing to do with inferences between propositions. The brain appears to be playing a different game from the game that [folk psychology] ascribes to it'.[31]

While Churchland has, in this paper, presented a case merely for the elimination of sententially-structured tokens in the brain, this last quotation demonstrates how he may be tempted to go beyond this claim, to the stronger claim that propositional description of our inner states is inappropriate (i.e. to make the transition from sentential- to intentional eliminativism, in Clark's terms). If the brain is 'playing a different game' from that postulated by folk psychology, then isn't this a compelling ground for eliminating descriptions using folk-psychological locutions? As Clark points out, Churchland succumbs to this temptation in the speculative concluding section of his paper 'Eliminative Materialism and the Propositional Attitudes'. Here Churchland contemplates 'three scenarios in which the operative conception of cognitive activity is progressively divorced from the forms and categories that characterise natural language'.[32] On the first of these scenarios, neuroscience discovers 'a new kinematics and correlative dynamics for what is now thought of as cognitive activity'.[33] What this amounts to is empirical support for sentential eliminativism - where what sentential eliminativism eliminates is both sententially-structured tokens in the brain, and a dynamics directly analogous to the relations of entailment etc. which obtain between the propositions which these sentence represent. Churchland is, on his own terms, charitable to folk psychology at this stage - folk-psychological accounts of the inner reality 'do carry significant information regarding it, and are thus fit to function as elements in a communication system'.[34] Folk psychological descriptions are, however, likened to the shadows on the wall of Plato's cave. The folk description is capturing only a 'narrow part of the reality' .Hence the temptation to eschew all use of folk psychology, at least in any attempt to convey accurately how cognition operates.[35]

This raises the question of how inaccurate a set of descriptive resources may be permitted to be, without succumbing to elimination. Fodorian enthusiasts for folk psychology will surely take any such description as *"(x)* believes that *(p)"* to be a partial and simplistic account of one part of the 'rational economy of propositional attitudes', which will embrace occurrent beliefs; dispositional beliefs; (perhaps) subconscious beliefs; beliefs which are held with greater and lesser degrees of conviction; beliefs which *(x)* might relinquish on closer inspection, in the light of discovered inconsistencies with other beliefs which *(x)* holds true, and so on. The difficulty with Churchland's account here is that it seems to depend upon a straw man argument, in that no proponent of folk psychology will take its descriptive resources to be generally entirely accurate and exhaustive in their account of an individual's cognitive state. It will, surely, be a largely arbitrary matter of stipulation as to how inaccurate a particular mode of description will be permitted to be, without being subject to elimination.[36]

As Churchland appears to suggest, his putatively superior descriptive apparatus is itself capable of rendering only an approximate account of what is going on during cognitive processes: he refers to '"solid" states' (quotation marks in original)[37], and concedes that 'exhaustive accounting of all dynamically relevant adjacent "solids" is not practically possible'.[38] So, prima facie, we have *two* partial accounts of the putative inner reality. Given that one of them - the folk account - is familiar and easy to apply (and enjoys considerable predictive success), it might be thought difficult to see what might motivate the difficult transition to relinquishing folk psychology's

descriptive apparatus, and replacing this with an account in terms of "solids" within a four- or five dimensional phase space.

There is, however, a stronger motivation than I have hitherto attributed to Churchland, and which becomes evident when he comments that '[folk psychological descriptions of cognitive processes are] ... unfit to represent the deeper reality in all its kinematically, dynamically, and even normatively relevant aspects'.[39] 'Folk theory-theory' is again performing a role in supporting eliminativism - hence the demand that folk psychology 'represent the deeper reality' .Having earlier been used in an attempt to justify the claim that folk psychology must reduce or be eliminated, the theory-theory thesis here compels the transition from sentential eliminativism to intentional eliminativism, so that *description* of mental states which avails itself of propositions and relations between propositions is to be rejected. As a theory, folk psychology carries with it a set of ontological commitments - and, as Churchland suggests in the opening paragraph of 'Eliminative Materialism and the Propositional Attitudes', '... both the principles and the ontology of that theory will eventually be displaced ... by completed neuroscience' .Given the theoretical status of folk psychology, there cannot be an employment of the theory which carries with it no commitment to the existence of token sentences in the head - as this is the ontology of the theory. If the use of folk psychological locutions entails that there are sentences in the head, then the claim that there are no sentences in the head entails the falsity of folk psychology qua theory.[40]

If the central claim of sentential eliminativism is false[41], then intentional eliminativism must be false: if there are tokens in the head which are tokens of propositional attitudes, then it cannot be erroneous to utilise intentional idioms to describe mental processes. But, as already noted, the truth of the sentential eliminativist claim is necessary, but not sufficient for intentional eliminativism. Dennett, for example, will endorse sentential eliminativism, but, as an instrumentalist, will not endorse intentional eliminativism - citing the falsity of the theory-theory thesis as the reason. Sentential eliminativism is, in fact, a very modest thesis - and certainly seems far too modest to accommodate the Kuhnian ambitions of Churchland.[42] Nor need the truth of the sentential eliminativist thesis entail the falsity of folk psychological attributions. If folk psychology is not a theory, then there are no corresponding theoretical entities - which would, on Churchland's account, be the mentalese tokens. If the claim that *(x)* believes that *(p)* does not entail the claim that *(x)* stands in a believing relationship to some physical brain token which represents that *(p)*, then the fact that *(x)* does not stand in any such physical relationship will not imperil the predictive and explanatory adequacy of folk psychology, which will survive whatever findings neuroscience delivers. This demonstrates the pressing need on the part of eliminativists for folk psychology to be taken to be a theory. For sentential eliminativism without theory-theory will not threaten 'the greatest intellectual catastrophe in the history of our species.'.[43]

But it seems that it might also be possible for sentential eliminativism's central claim to be held true, and for the theory-theory claim to be true, but for intentional eliminativism yet not to be vindicated. This would arise if folk psychology were, qua

theory, to quantify over entities which, *as conventionally described*, didn't exist[44] - but where its explanations and predictions retained their accuracy (as presumably they must, under these circumstances). The problem which arises here is the problem of stating under what circumstances an established theory's ontological commitments are false: as noted in my last chapter, this is the problem to which Stich alludes, when he asks:

> 'could it not be the case that folk psychology and computational theories are talking about exactly the same things, and that folk theory is just *wrong* about them?'.[45]

If, as Churchland claims, the entities being studied by connectionists have 'semantic' and 'syntactic' properties, then it is by no means clear that the appropriate 'judgement call' is in favour of the claim that folk psychology and neurophilosophy are talking about different things, rather than the claim that they are talking about the same things, but folk psychology is putatively mistaken in its characterisation of them. Churchland clearly needs to press the former claim in order for eliminativism to have any substance: as Clark argues, '... most recently published pro-Eliminativist writings manage to support [only sentential eliminativism]'.[46] If sentential eliminativism is to be construed as being the claim that folk psychology is wrong about some details of mentality, then, if that claim is taken to be equivalent to 'eliminative materialism', then surely no philosopher (or informed layman) will be anything *other* than an eliminative materialist.

The claim which sentential eliminativism contests is thus ill-defined, and hence sentential eliminativism is itself a vague thesis. Stich presents the contested thesis with a literal construal of the claim that there are 'sentences in the head', which are analogous to messages on a CRT:

> '... Fodor takes the objects of thought to be sentences in a species-wide mental code, the language of thought.'[47]

Horgan and Tienson, on the other hand, suggest that:

> 'a cognizer has a language of thought if it has a system of syntactically structured representations that undergo processes sensitive to that structure.'[48]

The fact of Churchland's using the terms 'semantics' and 'syntax' - albeit in scare quotes[49] - serves to demonstrate how elusive is the claim that sentential eliminativism seeks to oppose. Churchland uses the term 'representation' to describe the role of 'patterns of activation levels of neurons' - these giving rise to '... principled *transformation(s)*' .It is instructive to read this in light of Fodor's assessment of the conditions for the vindication of folk psychology:

> 'Common sense would be vindicated if some good theory of the mind proved to be committed to entities which - like the [propositional] attitudes - are both semantically evaluable and etiologically involved'.[50]

It's unclear what grounds - if any - Fodor would advance for not accepting that Churchland's activation vectors, if proven empirically to exist, and to 'represent' and 'transform', 'vindicate common sense' in the requisite manner. Part of the difficulty is,

of course, with the term 'representation': the claim that (a) represents (b) is, in this context, impossible to justify - for the reason advanced by Madell:

> 'No physical item can intrinsically be a representation of anything. As Dennett has argued, it can be a representation only *for* or *to* someone'.[51]

The insistently third-person approach adopted by physicalists such as Fodor and Churchland must raise a further problem, in that they cannot accommodate in their theses a self for whom the representation *is* a representation. Even if, per impossibile, there were to be some means of introspecting activation vectors amongst the neurons in one's head, the account offered by the physicalist would be mired in infinite regress, as awareness of this would be in terms of an activation vector, which would in turn be known only by a further activation vector ...

Of course, while Fodor may not take Churchland's speculations as being a potential vindication of folk psychology, Churchland will certainly not take his findings in this way. For him, the criterion for vindication would be theoretical integration of folk psychology qua theory - so that he must reject Fodor's criteria for vindication cited above. However, at the token level, what Churchland offers, and what Fodor claims to seek, seem to correspond. The difference is, I take it, in the nature of the aetiology offered by Churchland and sought by Fodor. Fodor wants causal mechanisms which correspond to the logical connections between propositions - and Churchland claims that empirical research work into activation vectors reveals causal connections which don't have this property. Thus Fodor:

> 'What connects the causal-historical aspect of [trains of thought] ... to its plausible- inference aspect is a general principle whose significance can hardly be over-emphasised: the train of thoughts that causes one to believe that such and such provides ... *reasonable grounds* for believing that such and such. Were this not the case - were there not this harmony between the semantic contents of thoughts and their causal powers - there wouldn't, after all, be much profit in thinking'.[52]

And, by contrast, Churchland:

> ... the most general and fundamental form of representation in the brain has nothing discernible to do with propositions, and the most general and fundamental form of computation in the brain has nothing discernible to do with inferences between propositions. The brain appears to be playing a different game from the game that [folk psychology] ascribes to it.[53]

Churchland's position here seems problematic in two respects: firstly, it is unclear what one would look for were one to look for physical features of the brain which *did* have 'something discernible to do with propositions', or whose causal transactions had something to do with inferences between these propositions. That what research evidence there is has found no appearance of the brain 'playing this game' seems hardly surprising. But, secondly, if the states which Churchland identifies *don't* play this game, then Churchland must surely have no adequate response to Fodor's point that the account which neuroscience offers in any putative science of mind must respect the inferences which obtain between propositional attitudes during instances of rational thought.[54]

Churchland thus faces a difficulty: the empirical evidence which he cites can, at best, support only the claim that there are no sentences in the head. Given the difficulty in proving the truth of a negative empirical existential statement, it may be doubted

whether even this much is likely to be demonstrable by connectionist-inspired empirical research. The further claim that folk psychology is a theory is not empirically demonstrable. But the claim that there are no sentences in the head will not, of itself, support the elimination of folk psychology, as outlined in my opening chapter. Even with the addition of the further claim that folk psychology is a theory, a 'judgement call' is required in order to press the further claim for its elimination, due to the difficulty in resolving the 'reference' problem raised by Stich and Lycan. But, having made this judgement call - on what can only be rather impressionistic grounds of 'gut feeling', Churchland now faces a new problem: his own use of propositional description of inner states notwithstanding, description of such states in these terms is in fact erroneous - and is presumably justified only on the grounds that we have not yet 'constructed a new [house] ... that invites us to move in.[55] But intentional eliminativism, while genuinely eliminative, and hence not open to the criticism of being too modest which I have levelled against sentential eliminativism, is now confronted with a problem: how can this position preserve our perception of ourselves as rational agents? Fodor's constraint on a science of mind - that it must respect the inferences which obtain between propositional attitudes during instances of rational thought - cannot be dismissed. Even if propositional attitudes are not physically tokened in the brain, even if folk psychology is a theory which has as its ontology such tokens, the fact remains that our understanding of rationality is ineliminably tied to behaviour being rational in virtue of its being consistent with the antecedent beliefs and desires of the agent. If intentional eliminativism is held true, then this self-image must be held to be false.

The eliminativist's options at this stage would appear to be:

i. deny the conclusion by pressing the more modest thesis of sentential eliminativism;

ii. accept the conclusion;

or

iii. attempt to postulate some basis for preserving rational agenthood in some form which is describable without recourse to the propositional-style locutions utilised by the folk.

Option (i) presents the difficulties to which I earlier alluded - that such a position will render the eliminativist thesis so modest that philosophers who are not even physicalists will endorse it, and that in consequence the eliminativist position cannot sustain Churchland's ambitions for a Kuhnian paradigm shift.

Option (ii) is surely incoherent. There can be no rational argument for the abandonment of rationality - which is equivalent to there being no argument for this position. Churchland may object that this is a purely a priori argument - and hence intrinsically antagonistic towards the epistemology which makes eliminativism possible, and that such paradoxes '... signal only the depth and far-reaching character of the conceptual revolution that [eliminative materialism] would have us contemplate'.[56] But this defence relies upon an epistemology which is in turn left undefended by Churchland.[57] Furthermore, the promise of a non-incoherent formulation is - even if we accept Churchland's epistemological stance - based on an

what is left as a remote future scientific possibility, so that Churchland's earlier-noted claim that the true task of the philosopher is speculation regarding the future discoveries of science is given an enormous burden of defence to perform on this occasion.

Option (iii) may thus seem attractive to the eliminativist, as an alternative to both weakness and incoherence. Given his epistemological stance, Churchland can feel free to reject the claim that it is analytic that agenthood entails rationality, and that rationality in turn entails the claim that rational procedures may be characterised propositionally. The task will be to envisage a radical reconfiguration of natural language - so that rationality can be sustained, but in some currently unimaginable form. This is the bizarre possibility which Churchland takes as the second of his 'scenarios in which the operative definition of cognitive activity is progressively divorced from the forms and categories that characterise natural language'.[58]

On this second scenario, we are:

> ... guided by our new understanding of those internal structures [to] ... construct a new system of communication entirely distinct from natural language ... the compounded strings of this new system ... are not evaluated as true or false, nor are the relations between them remotely analogous to the relations of entailment, etc., that hold between sentences.[59]

The latter observation raises the question of how such a scenario *could* be taken to be an alternative account of rationality: once again Churchland can fall back on his rejection of the claim that we can stipulate a priori that rationality must be accounted

for via 'the relations of entailment, etc., that hold between sentences' .On this occasion there seems to be no recourse to the claim that it is a matter for judgement whether the new concept is the same as the old concept, only differently construed, or whether there is the replacement of one concept by an entirely new and alternative concept. If this form of eliminativism - what Clark aptly terms 'catastrophic eliminativism' - is to address the objection levelled against intentional eliminativism that it fails to account for rationality, then it is rationality which must accounted for via catastrophic eliminativism, independently of any analytic consideration.

Churchland attempts to clarify his position in a more recent paper, where he observes that:

> the bottom line claim of the eliminative materialist ... is and always has been that *the content and character of our social practices in the domain of mutual perception, explanation, anticipation, and behavioural interaction are going to change, and change substantially, with the dawning of a truly adequate neuropsychology.* (italics in original)[60]

Churchland's claim here is ambiguous. He alludes to social practice - so that something more radical than eliminability at the level of scientific practice is envisaged. He also envisages elimination in practice - which is stronger than eliminability in principle - in that the former is sufficient for the latter, but not vice-versa. The 'judgement call' is still necessary, with regard to intentional eliminativism: will it be taken to be mistaken to *describe* the rationality of 'our social practices' using propositional attitudes? If Churchland is proposing merely that our understanding of activation vectors, together with our resources for their description, will gain with

future scientific progress, so that we *actually choose* to dispense with propositional-attitude description but *need not*, then the 'bottom line' claim of eliminativism amounts to the very modest claims: (a) that the language of thought thesis is false; and (b) we will in future have a richer understanding of the causal/ representational entities which *do* exist in our heads. Only when combined with the further claims: that folk psychology is a theory; and that all theories which are not reducible to the physical ought to be eliminated, do we have a thesis which is effectively eliminative. But as I have earlier argued, the 'theory-theory' claim is not defensible, and the 'unity of science' claim is a mere dogmatic assertion - so that, as it stands in this quotation, the 'bottom line' of eliminativism is surprisingly modest.

It is instructive to read Adrian Cussins' response to Hannan's target paper in the light of Churchland's 'bottom line'.[61] Cussins' central claim here is that Hannan's criticisms: ' ... *all* depend on a simple failure to understand what it is that eliminativism proposes to eliminate and what it is that eliminativism proposes to substitute'.[62] Cussins opens his paper with the claim that eliminativism's critics assume that 'the only kind of content is propositional or conceptual content'[63], and that the critics therefore beg the question against 'embodied representation'.

It isn't clear from this what it is that Cussins is taking his opponents to be assuming. The claim that 'all content is propositional' can be endorsed without making the further claim that there is a language of thought (i.e. we can claim that all content must be open to propositional description, so that in that sense it is 'propositional') .As I have earlier urged, only in conjunction with the theory-theory does the critic appear to

endorse an ontology of internally-coded propositions. But Cussins' claim raises a further question: what is it for a representation to be an 'embodied representation'? .It is not clear in virtue of what any such 'representation' *is* a representation. As earlier suggested, a necessary condition for a representation being so is that it is a representation to or for some observer. Thus, for example, Putnam presents a case where:

> An ant is crawling on a patch of sand. As it crawls, it traces a line in the sand. By pure chance the line that it traces curves and recrosses itself in such a way that it ends up looking like a recognisable caricature of Winston Churchill. Has the ant traced a picture of Winston Churchill, a picture that *depicts* Churchill?[64]

As Putnam goes on to suggest, we must conclude that the line is not "in itself" a representation of anything rather than anything else. There can be no *intrinsic* representation: the line looks like Churchill only *to us* - as *interpreters* of the putatively representational entity. It thus appears that it is not sufficient for *(a)* being a representation of *(b)* that *(a)* be embodied. A further difficulty arises when we ask: what is it for a representation to be 'embodied' ? I earlier noted that Churchland's response to Hannan's paper neglected directly to address the 'rationality' issue raised by the Pylyshyn 'burning building' example. It is unclear how 'embodied representation' can save rationality: the mere fact of the meanings of the various indications that the building is on fire being embodied within the head of the woman (whatever we must take this to mean) cannot be sufficient to account for the rationality of her fleeing the building. The representations thus encoded must be representations *to her* - and this latter component of any adequate account of rational choice seems likely to be missed by any such attempt to naturalise meaning.

Finally regarding Cussins' claim, we again encounter the problem of saying whether any putative account of rationality based upon embodied representation will be capable of *description* via propositional attitudes - that is, the question of whether Cussins is defending intentional eliminativism, or mere sentential eliminativism.

Cussins proceeds to set out three possible targets for elimination:

> (1) Folk psychological or common-sense psychological practice
> (2) Content or meaning or belief or rationality ...
> (3) The purported referents of Propositional Attitude theories of psychology'. (ellipsis in original)[65]

The first of these is a straw man argument, based upon a confusion between *eliminability in principle* - which Churchland and others do postulate - and 'the grotesque ... political proposal'[66] which no critic attributes to eliminativism, namely that individuals be somehow prohibited from using folk psychological locutions.[67]

With the second of these possible targets for elimination, we return to the issue of whether it is in fact possible for eliminativism to preserve meaning and rationality. This will depend in part upon the nature of the claim which eliminativism does in fact make. Cussins here makes the claim that 'meaning content, belief and rationality are in the world and likely to remain so'.[68] I take it that what is meant here is that in virtue of being embodied, these semantic properties are naturalised - 'in the world'. Cussins' attempted vindication of eliminativism depends upon keeping options (2) and (3) above separate - that is, he must maintain that one can eliminate sentences-in-the-head,

but nonetheless retain meaning and rationality. Interestingly, this is taken to entail the possible retention of representational content by the eliminativist:

> ... these consequences[69] would follow if eliminativism abandons representational content, but it does not ...[70]

What is at stake is here made quite explicit: if the abandonment of representational content has catastrophic, massively counterintuitive consequences, then the eliminativist must either embrace these consequences or retain representational content. But the retention of representational content surely impels the eliminativist towards a thesis so modest that it really does not merit the designation 'eliminativist':

> ... the *real* eliminativism - eliminates meanings-as-propositions and beliefs-as-propositional attitudes by rejecting the propositional *theory* of meaning and the propositional attitude *theory* of belief'.[71]

If, on Cussins' account, meaning (i.e. content) and belief is embodied, but is non-propositional, then all that his 'real eliminativism' eliminates is something akin to Fodor's language of thought. Cussins goes on to suggest that '... most of the arguments marshalled against eliminativism do not carry any weight against this third construal of eliminativism'. [72] This seems hardly surprising, as the third construal will find favour with such trenchant (and diverse) *anti*-eliminativists as Lynn Rudder Baker and Geoffrey Madell.

In his concern to render the eliminativist thesis moderate, and hence immune to the criticisms of allegedly misguided anti-eliminativists, Cussins even goes so far as to attack, on the Churchlands' behalf, intentional eliminativism:

> 'it is no part of the Churchlands' program to engage in the dubiously coherent exercise of rejecting all possible theories of representational content'.[73]

It is unclear what Cussins means by the term 'dubiously coherent' - but in deploying an argument frequently levelled by anti-eliminativists (i.e. that eliminativism, in denying the reality of belief, or truth or rationality, is self-defeating), Cussins appears to be endorsing traditional epistemological dichotomies - those which are abandoned by 'the epistemology that makes eliminativism possible' .For the naturalised epistemologist, the criterion for coherence will be coherence with the rest of science (philosophy differing from science merely by degree of entrenchment of the claims which it makes) - hence the case for elimination in the light of principled irreducibility. I leave to the following chapter the question of whether representation, together with the semantic network of which it is a key component, *can* be 'naturalised ' .

[1] Andy Clark, 'The Varieties of Eliminativism: Sentential, Intentional and Catastrophic', in *Mind & Language* vol. 8, no. 2, Summer 1993, p.223.
[2] Barbara Hannan, 'Don't Stop Believing: The Case Against Eliminative Materialism', in 'Don't Stop Believing: The Case Against Eliminative Materialism' in *Mind & Language* vol. 8, no. 2, Summer 1993, p.165. Hannan's is the target paper for this special forum issue on the topic of eliminativism.
[3] It is difficult to see how the rationality of behaviour might be captured *other than* via propositional description, but this is not of immediate importance here.
[4] Barbara Hannan, 'Don't Stop Believing', p.173. The Pylyshyn example is taken from his 'Cognitive Representation and the Process-Architecture Distinction' *(Behavioural and Brain Sciences* 3 (1980)).
[5] op. cit.
[6] Paul M. Churchland, 'Evaluating our Self-Conception', in *Mind & Language* vol. 8, no. 2, Summer 1993, p.212
[7] Barbara Hannan, 'Don't Stop Believing', p.173.
[8] Lynne Rudder Baker, 'Eliminativism and an Argument from Science', in *Mind and Language* vol. 8, no. 2, Summer 1993, p.180.
[9] Baker shares Hannan's opposition to eliminativism - but on the different grounds that folk psychological concepts do not require to be vindicated by science.
[10] Barbara Hannan, 'Don't Stop Believing', p.175.
[11] op. cit., p.174
[12] Although Hannan appears to be offering a hostage to fortune here, there are precedents for what I have here characterised as a limited reading of the claim that propositional attitudes exist. Thus Fodor: 'I propose to say that someone is a *Realist* about propositional attitudes if (a) he holds that there are mental states whose occurrences and interactions cause behaviour and do so, moreover, in ways that respect (at least to an approximation) the generalisations of common-sense belief/ desire psychology; and (b) he holds that these same causally efficacious mental states are also semantically evaluable' ('Fodor's Guide to Mental Representation' in Jerry A. Fodor, *A Theory of Content and Other Essays*, MIT Press, Cambridge Mass., 1990, p.5. There is scope for debate here as to the meanings of terms such as 'mental' and 'semantic' - but, as I will show later in this chapter, Paul Churchland attributes causal efficacy and the property of representation to 'activation vectors' - but Churchland would emphatically *not* regard himself as 'a realist about propositional attitudes' .Hence my claim that Hannan's proposal that 'propositional attitudes exist' is open to misconstrual by her critics, if it is intended that this claim amount to no more than the claim that there exist states which are causal and representational. I cannot conceive of any position in philosophy of mind which could coherently *deny* this.
[13] Lynne Rudder Baker, 'Eliminativism and an Argument from Science', p.182.
[14] Ch. 2
[15] Barbara Hannan, 'Don't Stop Believing', p.167.
[16] Hannan is concerned to reject the 'theory-theory' - a task which is made unnecessarily difficult by her equivocal support for 'post-logical positivist suspicions about the integrity of the analytic-synthetic distinction' ('Don't Stop Believing', p.167.).
[17] 'Don't Stop Believing', p.175.

[18] op. cit.
[19] op. cit., p173.
[20] Paul M. Churchland, 'Evaluating our Self-Conception', p.219.
[21] Churchland's apparent refusal to use the terminology of philosophical logic may be inspired by his own unequivocal commitment to 'the standards of epistemological evaluation that naturally go with [eliminative materialism]' (op cit. p.211) - but may also reveal some degree of endorsement of 'catastrophic eliminativism' - an especially radical form of eliminativism which proposes the elimination of all *use* of existing forms of natural language, and thus of the semantic and syntactic properties of natural language - including relations of entailment.
[22] Paul M. Churchland, 'Evaluating our Self-Conception', p.219.
[23] Paul M. Churchland, *Matter and Consciousness : A Contemporary Introduction to the Philosophy of Mind*, (Revised edition), MIT Press, Cambridge Massachusetts, 1988, p.24. I will consider in ch.5 Churchland's attempt to account for stimulus independence in terms of 'spontaneity'.
[24] Paul M. Churchland, 'Evaluating our Self-Conception', p.211.
[25] I take it that it is safe to assume that the expressed views of one of the Churchlands will be endorsed by the other, at least on such a substantive issue. In the preface to his most recent text: *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Mass., 1995,
Paul Churchland suggests that he and his wife "have become the left and right hemispheres of a single brain" (pxii).
[26] Patricia Smith Churchland, *Neurophilosophy*, MIT Press, Cambridge, Mass., 1986, p.283.
[27] Andy Clark, 'The Varieties of Eliminativism: Sentential, Intentional and Catastrophic', p.223.
[28] Paul M. Churchland, 'Evaluating our Self-Conception', p.220.
[29] op. cit.
[30] Paul M. Churchland, 'Explanation: A PDP Approach' in his *A Neurocomputational Perspective*, p.207. Interestingly, Churchland goes on to emphasise that the account which he has just given 'will appear very stimulus-responsish to many eyes', but that this construal would be 'oversimple and deeply misleading' - so that we have confirmation that Churchland would accept the first of my reconstituted Hannan premises.
[31] Paul M. Churchland, 'Evaluating our Self-Conception', p.221.
[32] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes, in William G. Lycan (ed.): *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Massachusetts, 1990, p.218.
[33] It is unclear whether this will, on Churchland's account, constitute a new concept of cognition, or an alternative account of how the existing concept is instantiated by an individual - but this seems not to be of concern for present purposes.
[34] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', p.219.
[35] I return to consideration of this part of Churchland's essay in my concluding chapter.
[36] A similarly arbitrary decision arises in Ramsey Stich and Garon's account of

the (related) choice between elimination and reduction of a theory. The authors conclude that 'since there is no easy measure of how "deeply and fundamentally different" a pair of posits are, *the conclusion we reach is bound to be a judgement call*' (Ramsey, W., Stich, S., and Garon, J., 'Connectionism, Eliminativism, and the Future of Folk Psychology' in John D. Greenwood (ed.), *The Future of Folk Psychology*, Cambridge University Press, Cambridge, Mass., 1991, p96.).

[37] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', p.218.

[38] op. cit., p219.

[39] op. cit.

[40] There is a further - though perhaps relatively unimportant - motivation behind Churchland's temptation to subscribe to intentional eliminativism. In his anthology, Lycan has separate sections of the text for 'classical' and 'current' eliminativism (William G. Lycan, *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Massachusetts, 1990). No account of the nature of the possible distinction between these two is offered in the editor's preamble; the use of the latter term may be taken to indicate that the distinction is merely temporal. The sole representative of classical eliminativism in Lycan's text is Feyerabend who, the reader is informed, ' ... was the first to argue openly that the mental categories of folk psychology simply fail to capture anything in physical reality and that everyday mental ascriptions are therefore false.' (op. cit. p.201). Feyerabend, then, appears here as a pioneer of the extreme 'explanatory adequacy of physics' thesis, whereby a new theory of cognition can - and ought to - be formulated 'without any recourse to existent terminology' .On Feyerabend's 'classical' account, intentional eliminativism is a goal which we *ought* to pursue. Churchland frequently acknowledges his intellectual debt to the classical eliminativists - so that there may, in addition to the theory-theory, be an additional motivation - an argument from authority.

[41] I return presently to the question of what exactly the central claim *is*.

[42] If the central thesis of sentential eliminativism is that there are no 'sentences in the head', and this suffices for one to be an eliminativist, then Geoffrey Madell is an eliminativist - and outcome which Madell will surely reject.

[43] Jerry A. Fodor, *Psychosemantics - The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge, Mass., 1987, p.xii. Fodor's language of thought will - if proven to exist - forestall this catastrophe, as its existence will entail the falsity of sentential eliminativism, and thus of intentional eliminativism, so that folk psychology is vindicated. It is interesting to note here that Fodor's 'catastrophe' would arise with intentional eliminativism - that is, before any slide from there into the more extreme variant of 'catastrophic eliminativism'.

[44] The problem here is, of course, that there *is no* 'convention' which can be falsified here. Folk psychology is not a theory, and hence incorporates no ontological claims - either implicitly or explicitly.

[45] Stephen Stich, 'What is a Theory of Mental Representation?', in Stephen Stich and Ted A. Warfield (eds.): *Mental Representation - A Reader*, Blackwell, Oxford, 1994, p.358.

[46] Andy Clark, 'The Varieties of Eliminativism: Sentential, Intentional and Catastrophic', p.227. Of course, as already suggested, the eliminativist commitment to

theory-theory will combine with this to yield intentional eliminativism, which is more radical. But in the absence of support for theory-theory, mere sentential eliminativism combined with a refusal to accept the instrumentalist position is required to yield intentional eliminativism - a position which Churchland himself violates by his frequent propositional references to, for example, knowledge states (e.g. the 'hickory nuts' example cited earlier).

[47] Stephen P. Stich, *From Folk Psychology to Cognitive Science - The Case Against Belief*, MIT Press, Cambridge, Mass., 1983, p.41.

[48] Terence Horgan and John Tienson, *Connectionism and the Philosophy of Psychology,* MIT Press, Cambridge, Mass., 1996, p.71.

[49] Paul M. Churchland, 'Evaluating Our Self-Conception', p.220.

[50] Jerry A. Fodor, *Psychosemantics*, p.26.

[51] Geoffrey Madell, *Mind and Materialism*, Edinburgh University Press, Edinburgh, 1988, p.27.

[52] Jerry A. Fodor, 'The Big Idea: Can There be a Science of Mind?', in *The Times Literary Supplement*, 3.7.92.

[53] Paul M. Churchland, 'Evaluating Our Self-Conception', p.221.

[54] I will later argue that problems in accounting for rationality may lead eliminativism into the hazardous territory of 'catastrophic eliminativism' where, it seems, rationality is itself eliminated.

[55] Paul M. Churchland, 'Evaluating Our Self-Conception', p.219. Here is the point: as the analogy with shadows on the wall of Plato's cave makes clear, folk psychology is not, it seems, *fundamentally* mistaken - it is merely too inaccurate to be countenanced by a future scientific theory of mind, hence the 'judgement call'

[56] Paul M. Churchland, 'Evaluating Our Self-Conception', p.214.

[57] Like Hannan, Churchland appears to assume that the popularity of this anti- a priori epistemology renders defence unnecessary, and truth assured.

[58] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', p.218.

[59] op. cit., p220.

[60] Paul and Patricia Churchland, 'Clark's Connectionist Defense of Folk Psychology', in Robert N. McCauley (ed.), *The Churchlands and their Critics*, Blackwell, Oxford, 1996, p.254.

[61] Adrian Cussins, 'Nonconceptual Content and the Elimination of Misconceived Composites!', in *Mind and Language* vol.8, no.2, Summer 1993, p.234.

[62] op. cit. p. 236.

[63] op. cit. p. 235.

[64] Hilary Putnam, *Reason Truth and History* Cambridge University Press, Cambridge, 1981, p.1.

[65] Adrian Cussins, 'Nonconceptual Content and the Elimination of Misconceived Composites!', p.234.

[66] op. cit.

[67] It is interesting that Cussins notes, apropos this 'grotesque' claim, that 'I know of no *sustained* attempt to defend such a construal of eliminativism' (op. cit., p.235 my emphasis).

[68] op. cit. p. 235.

[69] The 'catastrophic' consequences such as abandonment of the idea that there are persons.
[70] Adrian Cussins: 'Nonconceptual Content and the Elimination of Misconceived Composites!', p.239.
[71] op. cit., p.236.
[72] op. cit.
[73] op. cit. p. 238.

CHAPTER 4 WHAT STRENGTH OF ELIMINATIVISM DOES PDP ACCOMMODATE[1]?

My last chapter has argued that eliminativists have a dilemma: one version of the eliminativist claim is too weak to be taken seriously as elimination ('sentential eliminativism', which merely eliminates 'sentences in the head'); the stronger claim that sentential descriptions of mental states and processes ought to be eliminated leaves us without any way of accounting for human rationality, unless supplemented by some as-yet unavailable non-sentential system of verbal communication.

Paul Churchland's most recent essays have demonstrated his enthusiasm for the 'prototype activation' model of cognition - which yields:

> a novel and unorthodox conception of what cognition consists in [and which derives from] current research in cognitive neurobiology and from PDP models of brain function.[2]

The precise contribution of each of these scientific disciplines is not specified by Churchland explicitly: I take it that neurobiology provides information on the synaptic firings and synaptic connections in the human brain, and that PDP modelling in artificial intelligence yields an outcome which is held to be consistent with the neurobiology, in that the synaptic dynamics identified by neurobiology could be the medium for the cognitive dynamics postulated by PDP researchers. For present purposes, what is most significant about the PDP models is:

> their almost complete dissociation from the *sentential* or *propositional* conception of what knowledge consists in, and from the conception of human information processing as *rule-governed inference*. (italics in original)[3]

Churchland's use here of the term 'almost complete' is confusing - as a great deal hangs on the extent to which dissociation from sententialism is achieved by the model. As I will demonstrate, the model has no room in its ontology for sentences or propositions or their surrogates, so that sentential eliminativism obtains for the models. [4] But the claim that there are no 'sentences in the head' is, as earlier argued, in itself too weak to stand as a genuinely eliminative materialism. The significantly stronger (and genuinely eliminative) claim that sentential *descriptions* of mental states and processes ought not to be utilised is not entailed by the modest discovery of sentential eliminativism at the level of models of cognition. There is a further possibility: that PDP models, while not themselves employing sentential surrogates, nonetheless replicate human cognition which *does* employ sentential surrogates. It is unclear what further evidence would be required in order to arrive at the conclusion that absence of sentential entities in the model entails their absence in human mind/brains: perhaps if the research findings of neurobiologists demonstrated configurations of synaptic activity which were demonstrably isomorphic with the configurations of prototype-activation in PDP models, and hence hopelessly ill-suited to performing 'classical' cognitive processes, then the likelihood that sentential eliminativism was true of both would have to be conceded. Even then, however, one would require a further argument for intentional eliminativism - hence the intriguing uncertainty surrounding Churchland's use of the word 'almost' in the quotation.[5]

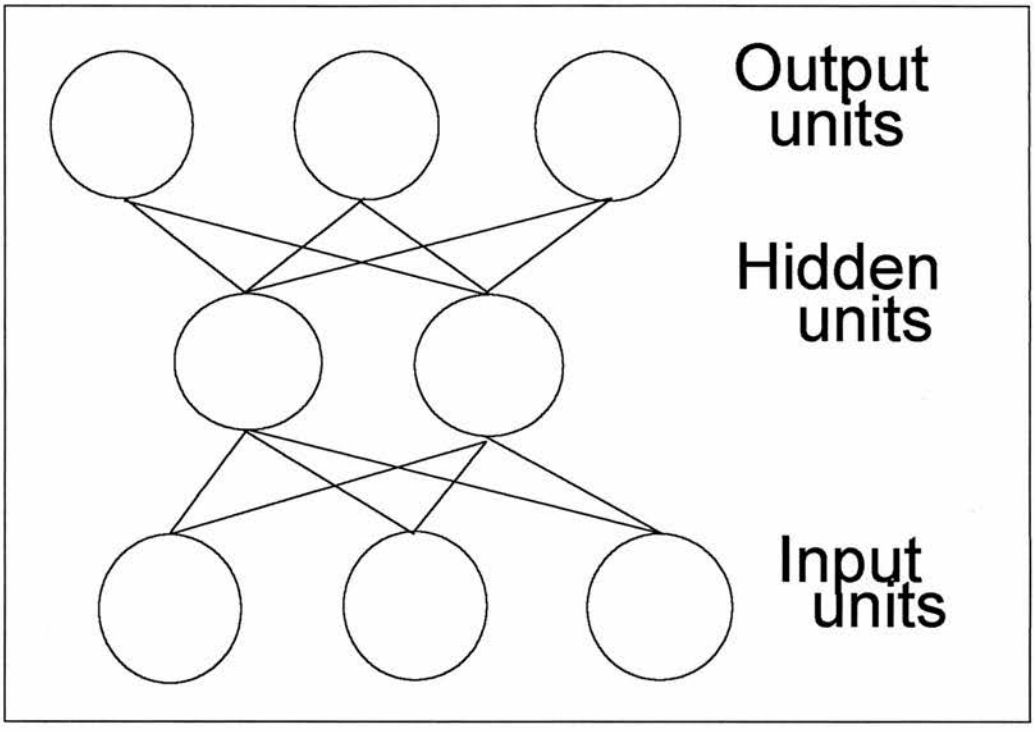
The prototype-activation model postulates no significant distinction between explanatory understanding (which is, for Churchland's present purposes, the explanandum), and perceptual recognition. His claim is that 'essentially the same kind of computational achievement underlies both...'[6]. This claim seems tendentious. If we say that to have an explanatory understanding is to be able to produce some account of how or why some phenomenon is as it is, then this is surely a more elaborate process than is perceptual recognition. It is explanatorily convenient for Churchland to postulate this close analogy - as the examples which he cites of actual successful PDP modelling (which I will shortly discuss) are examples of successful *recognition*. I can, however, recognise some *(x)* - such as a carburettor, or a van de graaff generator, without being able to do any more than state *what it is*. The 'computational achievement' involved in explanation is greater than this - even where my explanation is partial, or cursory, or prone to error in certain key respects. In his defence, Churchland will be able to claim, from this, that there are degrees of explanatory understanding. But there again seems to be an asymmetry between explanation and recognition, in that whether something is recognised as being a token of some particular type is an all-or-nothing affair.[7] Thus the fact that a model can putatively recognise does not entail that there is explanatory understanding, or the capacity for explanatory understanding.[8] The conflation of recognition and understanding is important, as Churchland goes on to present a PDP account of perceptual recognition - in the hope of having ipso facto produced a PDP account of explanatory understanding.

Churchland introduces his novel new account of cognition by alluding to the perceived shortcomings of its putative rival, the 'deductive-nomological' model of explanation. Churchland notes that:

> while much attention has been paid to the *logical* virtues and vices of this model, relatively little has been paid to its shortcomings when evaluated from a *psychological* point of view.[9]

This opening claim of Churchland's reveals a point which is crucial in his writing - namely, that for Churchland, theories and explanations are inside the mind/brains of those whose theories and explanations they are - and are not mere abstract entities. As a key figure in the development of the deductive-nomological model, Carl Hempel, in his *Aspects of Scientific Explanation*, did not appear to take himself to be doing psychology. Hempel's claim is that to give a scientific explanation of a phenomenon *(x)* is to show how *(x)* is subsumed under a law of nature. This can be set out in the form of a deductive argument - where the premises are the explanans, and the conclusion is the explanandum. *If* Hempel were to subscribe to the view that explanations are brain events, then, I take it, Hempel would endorse some model of mental processes such as Fodor's, with mental surrogates for each premise entering into causal relationships which eventuate in the mental surrogate for the sentence which is the conclusion of the argument. In this case, Hempel's model would, indeed, be a candidate for psychological criticism. If, on the other hand, Hempel is concerned exclusively with the logic, rather than the psychology, of explanation, then no Fodorian commitment is entailed by a thesis which does not locate explanations in the brains of those who formulate them.[10]

Churchland's alternative account derives from research which takes neurobiological research as being primary: neurobiologists present findings regarding the neuronal organisation of the brain, and artificial networks are subsequently developed which simulate or model the salient features of this neuronal organisation. What allegedly renders these artificial networks compelling is that they appear to be more psychologically realistic than more conventional, 'classical' models. One interesting example of this is the phenomenon of 'graceful degradation': interrupting a conventional, 'serial' program by taking out one tiny part will typically result in the destruction of the entire process; by contrast, the 'parallel' processing undertaken by connectionist networks can have whole sections removed, resulting in continuous degradation, rather than a sudden 'crash'. To the extent that human cognitive faculties such as recognition do degrade gracefully following some process such as the onset of dementia, the connectionist model is thus more effective in accounting for this than is the classical von Neumann serial-processing model.



The diagram above is an abstract depiction of the organisation of a PDP network.[11] There are three types of units: 'input units' are proximal to the external environment (e.g. the physical entity which is to be recognised, in the simple case of sensory recognition); 'output units' yield the ultimate result; the 'hidden units' (of which there may be a great many levels - only one is represented here) mediate between the input and output levels, it being here that the actual work of cognition takes place. For a functioning brain, Churchland's claim is that the input level neurons will be sensory neurons (so that in the case of visual perception, these will be located in the brain's visual cortex, and - as input units - these neurons will be the first to be activated by patterns of activity sent along the optic nerve). The lines in the diagram connecting each input unit to a hidden level unit represent synaptic connections in the brain (Churchland notes that in practice there may be hundreds of thousands of such connections from a single neuron, rather than the three here depicted). Via these synaptic connections, activation at the input level is transmitted to the hidden level(s).

This transmission may, for each separate connection, be either activation or inhibition - and to varying degrees, depending on the strength of the signal (the level of activation of the input unit), the strength of the synaptic connection, and its polarity (activational or inhibitory). Each hidden unit's activation is the aggregate of these three variables for each of the connected input units. The result is thus that a pattern of activation at the input level (an 'activation vector') will cause an activation vector at the hidden level. For any given activation vector at the input level, it is the configuration of synaptic weights which will determine the vector at the hidden level.[12] By the same process, the vector at the hidden level(s) result in an eventual output vector.

This model marks a departure from 'classical' models firstly in having no explicit propositional representations - that is, if the account of neural processing is correct, then sentential eliminativism is vindicated for human cognition, and not just for the models which have been constructed on PDP principles. Rather than being stored in some sentential surrogate, there is implicit and 'distributed' representation across an abstract and multidimensional 'vector space'. The pattern of connection, or activation vector, is the knowledge which the network possesses.[13] The 'parallel' element in PDP arises from the fact that neurons are actually quite slow in their operation - at around 100 firings per second. The greater 'psychological realism' which protagonists claim arises from the fact that many feats of recognition are both extremely complex, and virtually instantaneous. If facial recognition is achieved in, say, one second, then it must be achieved in 100 steps or less, as a result of the low firing speed. Given that the PDP model is a parallel processing model - with potentially enormous *vectors* of activity resulting from each synaptic firing - the model can putatively account for at

least some feats of human cognition which are difficult to account for using 'serial' (step-by-step) models.[14]

Churchland cites examples of the creation of artificial PDP networks which have successfully performed recognition tasks. The network's recognitional abilities are enhanced by the process of weight adjustment of the connections between the nodes at various levels, so as to manipulate the vectors which arise 'upstream' in consequence of any downstream vector. One example is the mine detector - which is to be trained to distinguish between the sonar echoes of mines and of rocks. This is achieved by providing the system with a variety of mine and rock echoes. Initially, the outputs will be unreliable (presumably yielding the correct answer - the output 'mine' following a mine echo input - approximately 50% of the time). The weights are then adjusted many times, until, for the initial training samples, the correct result is achieved 100% of the time. What is interesting in this case is that this 'training up' of the network generates a de facto theory of mine-and-rock echoes (the configuration of weights at the end of the training up) which cannot be specified in natural language.[15] Hence the theory is both intrinsic to the network (and not, *pace* Hempel, an abstract non-mental entity), and intrinsically non-sentential.

The immediate rejoinder is that this only serves to demonstrate the possibility of an artificial system for which sentential eliminativism is true - and that this is a considerably more modest outcome than the demonstration of intentional eliminativism at the level of artificial-network recognition - let alone at the level of human explanatory understanding (an outcome which Churchland requires[16]). If it is

fundamentally misleading (or even misleading to any significant degree) to say of the system that 'it has recognised a mine', then it isn't at all clear how this is misleading - so that intentional eliminativism is not sustained by this example.[17] But there are further problems which stand in the way of taking the mine-detector case to be conclusive. The vector codings which ultimately result from each input must be taken to represent 'mine echo' or 'rock echo', and, by analogy, vector codings in populations of neurons will be what mental representation consists in on the PDP model. But while the model is impressive as a model of computation, the other - representational - element in this model of human cognition is less compelling. Representations are only representations *to* or *for* some cognizer. The output of the submarine mine detector does not derive meaning from the environment which initially supplied the input: the output will take the form of some further representation, which is only a representation of 'mine' or 'rock' to the cognizer who is the machine operator on board the submarine. In much the same way as a pocket calculator doesn't present an output of numbers - but rather of numerals, which are *by human convention* invested with the particular representational significance which they have, similarly, the vector codings which emerge from a sequence of vector transformations in a connectionist device are intrinsically *bereft* of meaning.[18]

Churchland's theory of explanatory understanding has been described as employing prototype activation: it is during the training process that 'prototypes' emerge. At the hidden level(s), the training process progressively partitions the space of possible activation vectors, until, in the case of the mine detector, two approximate regions -

'hot spots' - emerge: one being the range of (hidden unit) vector codings for a prototypical mine; the other being the codings for a prototypical rock.

Having hitherto made use of cognitive neurobiology and PDP computational research, Churchland here accommodates a third area of current scientific endeavour: psychology. As Stich notes[19], Eleanor Rosch has presented a prototype theory of perceptual judgement as an alternative to 'assumptions underlying traditional philosophical analysis'.[20] These assumptions are those which underlie Socratic conceptual analysis, and the quest for necessary and sufficient conditions for the instantiation of a given concept. As Stich notes, the attempt to specify necessary and sufficient conditions for, for example, some *(x)* being beautiful, assumes that there *are* conditions such that if *(x)* satisfies them, then in consequence *(x)* must be beautiful, and conditions such that if *(x)* fails to satisfy these, then *(x)* cannot be beautiful. Such conditions will, of course, be generalizable. Stich notes that:

> there is now a fair amount of evidence suggesting that the assumptions underlying this traditional philosophical project may be simply mistaken. [21]

Rosch's research suggests that the mental structures which underlie our judgements when we classify items into categories 'do not employ tacitly known necessary and sufficient conditions for category membership, or anything roughly equivalent'.[22] What Rosch proposes as an alternative is a categorisation which:

> ... relies on prototypes, which may be thought of as idealised descriptions of the most typical or characteristic members of the category. The prototype for *bird*, for example, might include such features as flying, having feathers, singing, and a

> variety of others. In determining whether a particular instance falls within the category, subjects assess the *similarity* between the prototype and the instance being categorised. However, the features specified in the prototype are not even close to being necessary or sufficient conditions for membership. So, for example, an animal can lack one or many of the features of the prototypical bird, and still be classified as a bird. Emus are classified as birds though they can neither fly nor sing.[23]

Stich, and Rosch - and Churchland - have all noted the similarity between this model and Wittgenstein's 'family resemblance' theory of concepts. Among its many empirical advantages for Churchland is the fact that, if the training up process does indeed partition hidden vector space into prototype regions, then this will account for the already-noted advantage claimed for PDP models in their 'graceful degradation': the network may be able to render a correct verdict from degraded or partial input, as long as this input is sufficiently similar to a prototypical input. On the PDP model as presented by Churchland, the prototype region will pick out a 'family' (there is no single set of properties which all and only mine echoes possess - no necessary and sufficient conditions for some acoustic event being a mine echo - but the training up of the network results in the identification of a prototypical mine echo, and the system can then be relied upon, at the end of the training period, to identify non-standard mine echoes, as long as these possess sufficient of the properties of the mine prototype).

Churchland now moves from artificial networks to living creatures:

> the picture I am trying to evoke, of the cognitive lives of simple creatures, ascribes to them an organised 'library' of internal representations of various prototypical perceptual situations, situations to which prototypical *behaviours* are the computed output of the well-trained network..[24]

Churchland will later move, without further justification, from the 'simple creatures' to which he here alludes, to humans. But what is 'simple' in the model at this stage is the type of cognition which is being discussed - perceptual recognition - rather than the network whose cognition it is. There is, in any case, a need for some empirical justification for the analogy between artificial networks such as mine detectors and even 'simple' creatures. The reason for the inclusion of the 'simple creature' caveat soon becomes apparent, following Churchland's account of the operation of the prototype-activation model:

> the prototypical situations include feeding opportunities, grooming demands, territorial defence, predator avoidance, mating opportunities, offspring demands, and other similarly basic situations, to each of which a certain broad class of behaviours is appropriate. And within the various generic prototype representations at the appropriate level of hidden units, there will be subdivisions into more specific subprototypes whose activation prompts highly specific versions of the generic form of behaviour. (Is the mouse eating sesame seeds? Or hickory nuts? Is it avoiding a cat? Or a hawk?) These various prototypes are both united and distinguished by their relative positions in the hidden-unit vector space.[25]

Churchland's characterisation here of what is taking place within the creature's mind/brain ('Is it avoiding a cat? Or a hawk?) demonstrates how much additional work must be done in order to press the case for intentional eliminativism: the use of natural language and its sentential resources seems not only straightforward, but also *warranted* - so that only sentential eliminativism will be guaranteed by the accuracy of this account. Here we have input to the creature's sensory neurons initiating a vector which is transmitted to the hidden levels - where, I take it, prototypes and subprototypes will operate at distinct levels within a hierarchy of hidden levels. Behavioural prototypes are the outputs in this case - with the possibility of different

behaviours being triggered by different subprototypes (an animal which doesn't eat hickory nuts being prompted to no further action by the downstream activation of its 'hickory nut' subprototype; its eating sesame seeds, however, resulting in attack behaviour in response to the activation of its 'sesame seed' subprototype). The behavioural output can thus be rendered finely-grained by the partitioning of the hidden-level prototypes for 'food' into subprototypes for food which the creature in question eats, and food which it doesn't eat.

Churchland's motivation for the use of the qualifying term 'simple creature' is now apparent, as he notes that 'this picture will inevitably recall memories of behaviourism, for the perceptual environment is here portrayed as the fundamental control of motor behaviour, and the link between the two will appear very stimulus-responsish ... (sic)'. [26] Given that Churchland seeks to generalise beyond mine detectors and simple creatures performing recognitional tasks - ultimately to accommodate human moral action in his PDP model, the need to rebut suspicions of behaviour amounting to mere tropisms is clear. His response appeals merely to the simplicity of the heuristic model which has until now been utilised, and the correspondingly massively more complex architecture and functioning of the brain, which has:

> perhaps as many as a hundred layers along some pathways. Further, real brains divide into many distinct processing hierarchies working side by side on different problems. A brain is not a single network, but a committee of many co-operating networks - perhaps over a thousand of them in a typical mammalian brain. And most important for the present issue, the input to a given bank of hidden units comes not *just* from the sensory periphery, but from elsewhere in the brain itself. The brain is a *recurrent* network. The all-up input to any layer will almost always include some "current context" information that derives from earlier processing elsewhere in the brain.[27]

All that can safely be concluded from this is that *which* prototype or subprototype is activated will be a function of the sensory peripheral activity, *together with* any input from elsewhere in the brain - via projections from horizontally-adjacent layers, or downstream from levels closer to the output level. This possibility is presumably sanctioned by empirical findings in neurobiology - and is in principle capable of being designed into future artificial PDP networks. But the conclusion which Churchland draws from this disclosure of architectural complexity in living brains is far from being warranted: 'this frees us from the knee-jerk style of operation that worried us a few moments ago'.[28]

This conclusion is drawn on the grounds that:

> [the brain's] ultimate behaviour is a function of factors so many and so subtle, factors that interact in such highly volatile ways, that the brain's behaviour has become predictable only in its broadest outlines and only for very short periods into the future. Moreover, the factors controlling behaviour reside within the brain itself as much as in the external environment.[29]

There are thus two grounds for rejecting the 'stimulus-responsish' fear: firstly, complexity of the neural architecture of the human brain; and secondly, the internal origin of a significant proportion of those factors which eventuate in behaviour. Regarding the first of these - complexity - there seems to be a flaw in Churchland's argument, which utilises our ignorance in order to rebut the 'over-simple and deeply misleading' claim that the PDP model cannot accommodate human freedom. The nature of the argument here presented seems to be: we do not know all the factors

which govern behaviour; one of the factors may be free choice; therefore it is over-simple to claim that we have no free choice'. Clearly, one of the (currently mysterious) factors controlling behaviour 'within the brain' may account for freedom in some form - hence dispelling the 'knee-jerk' concern. However, it remains to be seen *how*, in principle, patterns of chemical activity across populations of neurons *can* achieve this. *Pace* Churchland's rather sanguine and cursory account here, predictability isn't the sole basis for claiming that the model is 'stimulus-responsish', and an argument from what may be no more than neurobiological speculation regarding internal processes of neural firings prior to instances of behaviour cannot, without much more work, be presented as our liberation from the fear that freedom can't be accommodated by the model.[30] This issue is important for my purposes, as it is still an open question how Churchland can respond to Pylyshyn's 'burning building' example, as presented by Hannan, and discussed in my last chapter. It will also be a consideration when we come, later in the thesis, to consider the question of what *strength* of eliminativism can be sanctioned by Churchland.[31]

In his *The Engine of Reason, the Seat of the Soul* - intended by Churchland to popularise neurophilosophy - he clearly recognises at the outset of the text how unappealing this aspect of his position might be to his audience. Once again, neural complexity is cited as ground for reassurance:

> ... your physical brain is far too complex and mercurial for its behaviour to be predicted in any but the broadest outlines or for any but the shortest distances into the future. Faced with the extraordinary dynamical features of a functioning brain, no device constructible in this universe could ever predict your behaviour, or your thoughts, with anything more than merely statistical success.[32]

It isn't clear what kind of impossibility is here being postulated: if it is empirically impossible for such a device to be constructed, then this is consistent with our having freedom of the will - so that there is a principled impossibility implied here. If on the other hand Churchland's point is that the human brain is just too complex for technology ever to reach the stage where such a device could be produced, then this is consistent with a principled possibility of predictability of human behaviour over the medium to long term with precision.[33]

Mere unpredictability thus may not satisfy Churchland's readers: freedom entails unpredictability - but unpredictability does not of itself entail freedom. All that can be deduced here is that between the initial stimulus and the ultimate response there is a highly complex internal sequence (even though, thanks to the enormous cognitive speed conferred by the parallel nature of the brain's operations, the actual time lag between stimulus and response is possibly so short as to give the false impression of there being no intervening influence on behaviour). This falls well short of a defence of freedom, which one would expect to find when reassured than an account is not 'stimulus-responsish'.[34]

When Churchland comes to account for the neural complexity which guarantees the absence of predictability, he introduces a new notion: 'limit cycles'. The massively recurrent nature of human cognition - whereby input can be horizontal and downward, in contrast to the purely upward input of the relatively much simpler mine detector - yields a situation where vectors cycle round a partially closed circuit within the brain,

eventually settling into a stable - and hence endlessly repeated - cycle, called a 'limit cycle':

> limit cycles are vital for orchestrating one's muscles to perform many familiar behaviours ... this is what keeps your heart muscle pumping ... breathing has a similar source. So do swimming, crawling, walking, running, flying, chewing, and almost every other repetitive or periodic behaviour ...[35]

One of the factors which will render my behaviour unpredictable is, surely, *my decision* to disrupt a limit cycle - for example, to hold my breath, as I might do while fleeing Pylyshyn's burning building. It is however difficult to see how the PDP model can accommodate this decision: it doesn't arrive as recurrent output from further upstream; nor is it an external input in the same way that physical stimuli from mines and rocks is an external input to the mine detector. It isn't the case that all input need be either recursive or sensory: Churchland alludes briefly to 'discursive' and 'linguistic' centres.[36] But such centres will receive and transform what has initially - at the input level - been physical stimuli. Churchland appears to have no room in his model for a free decision as an initiator of behaviour. Interestingly, Churchland concentrates on replacing only one element in the ontology of folk psychology - belief. It would be interesting to see how there can be a PDP surrogate for the other key ontological component, desire. My desire not to suffocate in the smoke-filled building, coupled with my decision to hold my breath as I run, interrupt the limit cycle which normally ensures that I keep breathing. This is an instance of rational behaviour, of course - and our conception of ourselves as rational beings is bracketed by Churchland, pending future developments in neuroscience. But my inclination is to resist such bracketing, and insist on the reality of such rational free choice.

Churchland recognises that unpredictability does not entail freedom of the will - but in doing so, he raises an interesting new problem, in his evident desire to render his account appealing:

> it would be foolish to mistake such (genuine) unpredictability for what philosophers and theologians have often hoped for in the way of free will. That term was typically meant to apply to a human capacity that *transcended* the natural order, whereas the dynamical picture here presented keeps us firmly embedded within the causal order. But it is legitimate to see it as the ground of something extremely important: our capacity, at least occasionally, for genuinely spontaneous activity...[37]

I take it that 'spontaneous' here means 'originating from internal or self-motivation'. Having made the point, Churchland promptly drops it without further elaboration. It is completely unclear what difference there is intended to be between the claim that my decision to hold my breath is 'free', and the claim that the decision is 'spontaneous'. Interestingly, the American Webster's Dictionary cites, at the opening of its entry for 'spontaneous', the Latin origin of the term: *'sponte:* of one's free will, voluntarily'. [38]

Other entries in Webster's for 'spontaneous' include: 'without external constraint'; 'unpremeditated'; 'self-acting'; 'indigenous'; and 'occurring in the natural course of things'. It isn't clear which, if any, of these Churchland could avail himself of without complication. If we ask the question whether artificial PDP systems can act 'spontaneously', I take it that the answer is negative. All that is lacking in the simpler networks is recurrent processing (as noted earlier, the mine detector's processing -

both inputs while operating, and 'backpropagation' during training up - is exclusively upstream, in contrast to the horizontal and downstream capacities of the human PDP system).[39] In the absence of any further significant distinction between artificial and human networks, recurrence must presumably be the basis for 'spontaneity'. The suspicion must be that Churchland is playing to the gallery on this occasion: all that *can* plausibly be meant by 'spontaneously' is 'purely internal' - though the non-philosophical audience for whom this text is primarily intended may assume that something superior to the free will fruitlessly pursued by 'philosophers and theologians' is on offer - 'genuinely spontaneous activity'. Even then, purely internal initiation is difficult to accommodate within the model as presented by Churchland, as already noted. The origin of a putatively 'spontaneous' action must be an antecedent physical event; this would appear to rule out freedom of the will - but, given the exclusively external and recurrent sources of all the processing to which Churchland alludes, the putatively purely internal physical event which initiates spontaneity is deeply mysterious.

The attempt to construct a PDP response to the challenge posed by Pylyshyn's 'burning building' serves to illustrate the explanatory gaps which are yet to be filled. In the first instance, the smell of smoke in the building - the physical event of the inhalation of smoke - will trigger input neurons upstream of the nostrils, in an olfactory region of abstract vector space. The vector of neural activity at this input level will then be transmitted upstream to the hidden levels units, for interpretation - allocation to some prototype at that level or levels. For the eventual response - fleeing the building - not to be a mere tropism, the vector at one of these levels must represent

"danger".[40] The output in this case is activity - running from the building - which will be the consequence of muscular responses to motor-neuronal outputs. But the smell of smoke at a barbecue, for example, does not eventuate in escape behaviour - so that the smell of smoke must *have meaning* in this context which it doesn't in other contexts. Churchland has insisted that a given vector of itself represents nothing - it represents only in the context of the global configuration of synaptic weights which produces it, in the light of incoming data. But this doesn't resolve the problem of how a pattern of chemical activity in the brain could have any meaning at all. Not only must the activation vector + synaptic weights have intrinsic meaning in order for the connection with a 'danger' prototype to be effected; in order to avoid the conclusion that the action is merely a tropism, the combinations of such vectors + weights must mean something to the individual whose brain it is.[41] Churchland may be able to offer some account of how enormously complex recurrent activity is taking place in the brain, so that prototypical smoke recognition is associated with a "danger" prototype elsewhere - this latter vectorial activity being the initiator of the motor output which manifests itself as running away. But such an explanation would still have to account for this association having been made in the first place (while it would not have been initiated at a barbecue). Again a purely causal explanation misses what must be a semantic property which accounts for the danger-association being initiated.

There is a second feature which the network must possess, in order to obviate the 'stimulus-responsish' concern arising vis-à-vis the burning building: as already noted, my choice not to be burnt or suffocated - or my choice not first to scour the building in search of others who need my assistance - must be recognised and accounted for. As

already urged, a pattern of [stimulus - very complex internal processing at very high speed - response] is inadequate as a reassurance, where the concern is that my escape behaviour is a mere tropism. In short, problems both of intentionality and of free will contrive to present enormous obstacles to the acceptance of any immediately-obvious PDP account of the escapee's behaviour. Given the prominence which Hannan accords the 'burning building' case in what is a target article to which Churchland has been invited to respond, his reticence in presenting his own account must be taken to be significant.

Setting aside for now any concerns regarding the potential of his PDP model to accommodate any account of human freedom, we can now summarise Churchland's alternative to the Deductive-Nomological model (which, as already noted, is taken by Churchland to be a model of the operation of the mind/brain). Whereas the D-N model has sentential structures as representations, with computation consisting in inferences made according to structure-sensitive rules, the PDP model has vector codings for representations, and computation is vector-to-vector transformation within the network. A theory is, on the PDP model, a configuration of synaptic weights - so that a single input vector will eventuate in different outputs, depending upon the theory which is applied to it. Churchland notes in passing that 'of course, the vectors themselves represent nothing, save in the context of the global configuration of synaptic weights that produced them'.[42] So prototypes arise *modulo some theory* - the vector in question representing what it does only in the context of the synaptic weights (i.e. theory) from which it eventuates. This will have the interesting result of making prototypes theoretical entities (theories being, as already noted, 'in the head').

Churchland's main claim vis-à-vis PDP networks in the essay 'Explanation: A PDP Approach' is that explanatory understanding may be accounted for in terms of prototype activation:

> ... explanatory understanding consists in the activation of a specific prototype vector in a well-trained network. It consists in the apprehension of the problematic case as an instance of a general type, *a type for which the creature has a detailed and well-informed representation.*[43]

Here again we have the appearance of sentential, but not intentional eliminativism: 'detail', and the property of being 'well-informed' seem unavoidably semantic - so that describing the network (or its possessor) as 'understanding that remaining in the burning building is dangerous' would appear not to be inappropriate. It is still unclear *how* this understanding arises, however: the 'problematic case' in question - that of the building's being on fire - needs to be apprehended under some prototype which represents 'danger' and the urgent need for escape, but given the apparent paucity of the input which may in practice trigger this realisation, accounting for the performance of the appropriate behaviour is both problematic, and, if the model is to be deemed plausible, urgent. This is reminiscent of the project identified by one of Churchland's mentors, Quine:

> the relation between the meagre input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence.[44]

The relation of evidence to theory - the process whereby the smell of smoke triggers activation of 'danger' prototypes - can apparently be accounted for only partially, as the consequence of massively recurrent activity, achieved virtually instantaneously due to the parallel nature of the processing involved, and the prior training of the network. What is missing is an explanation of the semantics involved: even though Churchland may wish to eschew semantic talk, in order to distance himself from classical sententialist accounts such as that of Fodor, there must nonetheless be some neural surrogate for semantics, in order that the significance of the smell of burning in this particular context can lead to the transition from a 'smoke recognition' vector to a 'danger recognition' vector.

There is a further criticism which can be levelled against this account, and it is a criticism of which Churchland is aware:

> "What you have outlined", runs the objection, "may be a successful account of spontaneous *classification*, but explanatory *understanding* surely involves a great deal more than mere classification.[45]

Clearly there is some work to be done here: the mine detector, following the completion of its training up, successfully classifies mines, by triggering its 'mine' prototype in response to incoming sonar echoes from a mine, and generating the appropriate output. But the mine detector does not have any explanatory understanding. Churchland's case thus far has been based largely on analogy between artificial networks such as the mine detector and human brains; the only significant difference between them (I exclude the material composition as being not significant)

is the capacity of the human brain for recurrent and horizontal processing. Unless there are other significant architectural distinctions, this greater complexity of processing must account for the human's possession, and the artificial network's lack of, explanatory understanding.[46]

Churchland's response to the objection which he has himself raised alludes to the contrast between meagre input and (relatively) torrential output:

> what we must remember is that the prototype vector embodies an enormous amount of information ... that vector has structure, whose function is to represent an overall *syndrome* of objective features, relations, sequences, and uniformities. Its activation by a given perceptual or other cognitive circumstance does not represent a loss of information. On the contrary, it represents a major and speculative *gain* in information, since the portrait it embodies typically goes far beyond the local and perspectivally limited information that may activate it on a given occasion. That is why the process is so useful: it is quite dramatically ampliative.[47]

This seems to miss the point of the objection which Churchland has himself raised. What we have here is more elaborate classification in consequence of the activations, further upstream, of progressively richer and more detailed prototypes. But it does still appear on this account to be classification, rather than explanatory understanding which is being initiated. The greater detail would be *an aid to* correspondingly greater understanding - but cannot of itself be taken to be what explanatory understanding consists in. What may be the source of my confusion here is Churchland's need to eschew traditional and sententialist accounts of what explanatory understanding consists in: if we are to say that to explain is to be able to answer a 'why' question, then this would appear prima facie to be uncongenial to an account which at a

minimum wishes to eliminate 'sentences in the head' - and, as previously argued, must go beyond this in order to sustain a genuinely eliminativist position which will meet the ambitions of the Churchlands. But a non-sententialist account of explanatory understanding cannot consist merely in an account of the progressive enrichment of classification via prototype: *pace* Churchland, we cannot accept that: 'on each such occasion, the creature ends up understanding ... far more about the explanandum situation than was strictly presented in the explanandum itself' without an account of how the transition from recognition to understanding is effected.[48] In the absence of further reassurance on this point - and further detail - the 'stimulus-responsish' objection again rears its head: the mine detector doesn't attempt to escape on detecting a mine in the vicinity; the woman in the building *does* try to escape on detecting smoke. But the difference resides, not in her understanding, and the machine's failing to understand, the danger - but in the greater sophistication of prototype in the woman's brain. This will account for the activation of motor-neurons (and hence her running away) via a purely causal story - but does not resolve the 'S-R' problem.

Churchland's development of the claim that explanatory understanding is via prototype activation raises some interesting issues. He argues, for example, that 'explanatory understanding is the same thing in all ... cases: what differs is the character of the prototype involved'.[49] This will lead to what I will call Churchland's 'continuity thesis': the claim that there is no fundamental difference between moral or social explanation and scientific explanation; these consist in the activation of different types of prototype vector - but this difference is not fundamental. Churchland cites various

examples of prototypes in order to illustrate the putative prototypical diversity of a well-trained human brain. The gap between classification and explanatory understanding is, however glossed over: the first category to be discussed is 'property-cluster prototypes' - which comprise the majority of our human prototypical vector space. But the 'explanatory role' is left mysterious by Churchland: '"Why is its neck so long, Daddy?" "It's a *swan* dear; swans have very long necks"'.[50]

This 'explanation' of the length of a swan's neck seems hopelessly impoverished - though it might well satisfy a small child (both of the examples cited by Churchland are of such responses to a child's hypothetical questioning). As an account of how parents respond to such questions this may well suffice - but it appears to be little more than further classification (the set of all things which are swans is a subset of the set of all things which have long necks). A proper explanation of *why* swans have long necks would surely go beyond mere classification - citing, presumably, the causal effects of evolutionary development on the species. This seems ineluctably sentential - so that if the sentential is to be expunged from the model, it is incumbent on Churchland to demonstrate how this is to be achieved, while still retaining what is recognisably explanatory understanding.

Churchland's next category of prototypes is 'etiological prototypes'. Again the account is one of recognition - an etiological prototype '*depicts* a typical temporal sequence of event types' (emphasis mine).[51] Churchland here appears to shift his position somewhat: in place of the *identification* of explanatory understanding and prototype activation, we here have a sequence which *'make[s] possible* our

explanatory understanding of the temporally extended world' (emphasis mine).[52] While Churchland denies that it is his intention to give an analytical definition of 'cause' - what intricacies constitute a *genuine* etiological prototype from a pseudo-etiological prototype is dismissed as a 'secondary [matter] I shall leave for a future occasion'[53] - the PDP model interestingly appears neatly to account for Hume's psychological contribution to causation: having made the neural connections which classify a given temporal event as a causal event, and having repeated this classification on a number of occasions, the classification may eventually become so habitual as to resemble a 'limit cycle'. Thus, while the model does not explain causation, it may be the basis for an explanation of how Hume's account of causation might operate in practice.

Given his philosophical commitments, Churchland is clearly concerned to illustrate how the process of competition between theories - and the subsequent reduction and elimination of theories - can be accounted for using the PDP model. Again the absence of distinction between explanation and recognition - and thus with mere classification - poses problems for Churchland's account of scientific progress. This lack of distinctiveness is made explicit in Churchland's central claim vis-à-vis scientific progress:

> the central phenomenon to be explored here is the brain's *vector completion* of partial or degraded inputs, a completion often aided by the brain's recurrent manipulation of the relevant population of representing neurons. In plain English, it is the phenomenon of your recognising - perhaps slowly at first, but then suddenly - some unfamiliar, puzzling, or otherwise problematic situation as being an instance or example of something well known to you.[54]

Churchland has previously prepared the ground for this claim, by giving examples of degraded images - photographs photocopied dozens of times, so that the image is obscured, for example - and demonstrating how we can nonetheless accurately classify the object which is only partially represented by the degraded image. The recognition of a problematic situation as falling under some prototype, hitherto not employed while considering situations of this type, bears a clear resemblance to Kuhn's 'paradigm shifts'.[55] The scientist is, on this account, utilising the enormous recurrent potential of the brain 'to explore a range of different activational possibilities'.[56] These alternative prototypes are already located in abstract vector space: scientific progress consists in the accommodation of the problematic situation under a novel prototype. This raises two immediate problems: (i) we need an account of what it is, in virtue of which the novel prototype is more appropriate than that utilised hitherto - that is, we need to know in virtue of what we have *progress* rather than mere novelty; and (ii) once again, the problem of accounting for the initiation of this process arises. Churchland's account - the utilisation by the scientist of novel prototypes, via his access to recurrent pathways - doesn't sound any different from the sonar operator's use of the mine detector on board the submarine. In his partial defence, Churchland will have difficulty in eschewing such dualist-sounding formulations - comparable to his difficulty in eschewing sententialist-sounding formulations - and we must avoid reading actual philosophical commitment into such formulations. But there must be a concern that Churchland is conflating the *running of a program* and the *instantiation of a program*: the fact that, for example, artificial PDP networks may be created which mimic the human capacity for phoneme-detection, does not entail that the mind/ brain is itself a PDP network. While

the mine- or phoneme-detector must be operated - inputs provided, output interpreted etc. - by an operator who is not the mine- or phoneme-detector, in the case of the scientist whose brain is putatively a natural PDP network, the network is - on Churchland's account - *operating itself*. This cannot be explained merely by recourse to the observation that the various parts of the brain interact, so that one part is accessing the resources or output of another part: a further account, privileging one part for at least some types of cognitive operation, and explaining the basis for this privileging - so that this part of the brain is the equivalent of the machine operator - is necessary. Otherwise the suspicion remains that what the artificial phoneme detector does is to run a program which *mimics* the operation of the human mind/ brain; in consequence, the presumed analogy lacks persuasiveness, and we thus have no grounds for accepting even sentential eliminativism other than for the machine.

In the absence of a sustained response by Churchland to this latter concern, I propose to proceed to consider how he might address the former concern: the question of what makes for progress, where progress consists in the deployment of a prototype in a novel way. Churchland takes the example of developments in cosmology to illustrate this in his 1995 text. His account opens rather obscurely, with primitive mankind viewing the 'degraded perceptual input' of the stars 'scattered carelessly both in space and in brightness'.[57] It isn't clear to me that this is an example of 'degraded perceptual input' comparable to the photocopier examples earlier cited - but in any case, the attribution of names to the constellations such as 'the dipper' is taken to be an attempt to 'impose a structural order on the contents of the night sky'.[58] If it

isn't clear what Churchland's point is here, the conclusion which he reaches is really eccentric:

> few of these interpretations are very compelling, visually. And certainly none of them yielded any useful predictions of stellar behaviour ... the scorpion never stung anything; the dipper never poured out any water. In this respect, these interpretations of the visual chaos were not "good theories" about stellar phenomena.[59]

This example is useful in the first instance for demonstrating how primitive humans pursue the task of apprehending perceived phenomena under some prototype. Their relative intellectual impoverishment manifests itself in the fact that they have far fewer recurrent pathways than do we - so that the range of prototypes immediately available for the task of 'imposing structural order' on the night sky is impoverished by our later standards. But this observation of Churchland's is combined with an indication of one of the tasks of theory - 'yielding useful predictions'. This raises two points: (i) once again, we have the gap between classification and, in this case, prediction rather than explanation. These cognitive tasks - classification and prediction - cannot be taken to be identical. Classification will be, at best, an aid to successful prediction; and (ii) this point is surely reinforced by the absurd suggestion that the failure of 'the dipper' to pour water demonstrates how poor was the theory. Clearly predictive adequacy is a bona fide criterion for theoretical virtue - but to suggest that this was the motivation behind the original naming of the constellations is to push the model beyond credibility. Churchland has previously argued - in my view unsuccessfully - that all explanatory understanding is prototype activation.[60] Here we have the quite separate - and surely implausible - claim that all prototype activation is explanatory

understanding. Churchland will understandably wish to distinguish between mere classification and explanation/ prediction[61] - but he cannot do so by eliminating classification; classification would surely on his account be prototype activation (I cannot see what else it might be), and that (i.e. classification) is what is occurring in the case of the primitive denomination of the constellations.

Progress takes place, on Churchland's continuing account, when the Greeks deploy the prototype 'rotating sphere'. This is a prototype presumably available to the primitives - but its application with a view to 'imposing structural order on the night sky' is novel. Churchland here reinforces the uncontroversial claim that predictive accuracy is a criterion for theoretical success: the motions of the stars and their future positions is now, via the 'rotating sphere' prototype, predictable with great accuracy. To make this point is to demonstrate the fatuity of the contrast with the primitive nomenclature - which not only had not had the intention of making this prediction, but probably had no predictive purpose at all. Perhaps surprisingly, Descartes is given credit by Churchland for the next stage in progress - with his activation of the 'vortex' prototype to account for the dynamics of the solar system. Churchland's description of the next stage in scientific progress is revealing:

> and once again, the interpretation was false. Or, at least, Sir Isaac Newton came up with a much better one.[62]

It isn't entirely clear how to read this observation: the phrase 'or at least' seems to cast doubt on the attribution of falsity to the Cartesian interpretation (this is repeated at the next stage, where Churchland notes, apropos Newton's interpretation, that 'this

brilliant interpretation eventually proved false as well. *Or at any rate*, Albert Einstein came up with a still better one' (emphasis mine).[63] Churchland's commitment to 'the epistemology that makes eliminative materialism possible'[64] has led him to bracket the notions of truth and falsity, it being unclear what might be the criteria - or whether any criteria exist - for attributing truth. Here, however, he seems to hedge his bets by twice appealing to falsity, only to enter an immediate caveat to the effect that the deployment of one prototype leads to (or, on Churchland's account, *is*) merely a better explanatory understanding of the explanandum. Churchland concludes this brief historical account:

> the point of this brief and highly selective excursion into the history of science has been to portray some of the most sophisticated of our intellectual achievements as involving the very same activities of vector processing, recurrent manipulation, prototype activation, *and prototype evaluation* as can be found in some of the simplest of our cognitive activities, such as recognising a dog in a low-grade photograph .(my emphasis)[65]

Aside from the reiteration of the continuity between the cognitive tasks of perceptual recognition and theory-formulation, the interesting observation here is the possibility of 'prototype evaluation'. Given that a prototype is an abstract region in vector space, its evaluation in advance of its translation into some suitably semantic form is hard to envisage - and, of course, the possibility of such translation is still, at this stage, moot. [66]

Given his doctrinal commitments, Churchland cannot not have a view on the possibility of theory evaluation, as his sometimes scathing comments on folk psychology demonstrate. It is therefore intriguing for the sceptic to come across, in

'Explanation: A PDP Approach', a sub section headed 'Inference to the Best Explanation': a subsection which opens with the words:

> the idea of prototype activation throws some much-needed light on the popular idea of "inference to the best explanation".[67]

Churchland poses the problem as posed by C.S. Pierce: 'for any set of observations there is a literal infinity of possible hypotheses that might be posed in explanation'.[68] Consistent with his approach to Hempelian explanation, Churchland reconceives the problem, presenting it not as a logical problem (as Pierce did), but, rather, as a psychological problem. "Inference to the best explanation" is, we are told, a 'crude notion' - to be here replaced by 'the more penetrating notion of "activation of the most penetrating activation vector"'.[69] The term 'more penetrating' is left vague - but Churchland notes that it is a psychological fact that, within the domain of psychological possibilities (as opposed to Pierce's logical possibilities) there just isn't the range of alternative hypotheses postulated by Pierce.[70]

Having recast Pierce's problem in psychological terms, Churchland then purports to resolve the problem:

> we do not search an infinite space of possible explanations. In general, we do not search at all: in familiar cases a suitable prototype is activated directly.[71]

Pierce's problem seems here to be ignored rather than resolved - and in a manner which again highlights the difficulty in presenting a plausible non-dualistic account: where a successful prototype vector fails to be activated directly, 'one repeatedly

reenters the problematic input', until a suitable activation is achieved. This relationship - between the controller of the process and the medium in which the process is conducted - is left by Churchland as if in no need of further clarification.

More seriously for my immediate concerns - which are epistemological rather than metaphysical - Churchland goes on to note that:

> the range of concurrently possible understandings is closed under the relation "is at least within hailing distance of an existing prototype".[72]

The issue of the relationship between different prototypes is an important one. As I have earlier suggested, Churchland needs some neural (i.e. causal, non-semantic) account of how, in the case of the woman in the burning building, there is a connection effected between the antecedent 'smoke' prototype, and the subsequent 'danger' prototype, to account for her action in running from the building. Similarly, I take it that the creative genius of Newton, in his novel application of the 'deflecting force' prototype to account for the moon's elliptical orbit, is to be accounted for - in terms of cognitive neurodynamics - by his entering the problematic input in a prototype vector distant from that being deployed by his relatively benighted predecessors.[73] But the problem now arises of accounting for the relationship 'being close to/ distant from', as applied to vectors. The vectors, qua vectors, exist in abstract space - so that the term 'proximity' must be purely figurative at this level. The respective activations of the various vectors will - qua brain events - occur in different regions of the brain, but it isn't clear that this literal construal of 'proximity' is what Churchland has in mind. The processes of horizontal and recurrent activity - or at least the horizontal element - may

be facilitated where there is a relatively small distance between the parts of the brain responsible for the prototype activations in question, but Churchland hasn't, to my knowledge, discussed this possibility, so that we cannot assume that the 'distance' is a literal distance.[74] A third, and again figurative, possible construal of the notion of 'proximity' in this context might be that it is proximity in the sense of semantic similarity. Thus, for example, the claims that the earth is a flat disc, and that the earth is a sphere, are, in some manner which may require disambiguation, 'semantically close' - the meanings of 'flat disc' and 'sphere' are distinct, but nonetheless similar; the claim that the earth is an idea in the mind of God, on the other hand, is 'distant' from either.[75]

The confusion engendered here is ironic, given Churchland's clarificatory purpose: 'closeness' between vectors is to account for one of the key elements in his account of scientific progress - analogy:

> a prototype vector whose activation has hitherto been confined to one empirical domain subsequently comes to be activated with profit in a new domain. More accurately, the new domain activates a vector that is *close* to the old prototype ... talk of analogy has always been hobbled by our inability to say anything very specific about what constitutes the relevant kind of similarity. We are now in a position to be entirely specific. Analogy ... consists in the close proximity of the respective prototype representations ... in the relevant hidden-unit activation-vector space. (emphasis in original)[76]

If Churchland glosses over the semantic problem with regard to analogy, he addresses it directly when he comes to recognise that the term 'best' in 'inference to the best explanation' is an evaluative term. When considering the question of what makes one prototype vector activation 'better' than another, 'we must answer carefully, since we

are denied the usual semantic vocabulary of reference, truth, consistency, entailment, and so forth'.[77] Here Churchland adopts an intentional eliminativist stance: were he merely to be advocating the elimination of 'sentences in the head', then, of course, the neural ontology with which these are to be replaced could stand proxy for sentences outside the head - and thus be candidates for consideration in these semantic terms. The 'best' explanation - and, indeed explanation itself - could then be that which is most successful when considered, using conventional semantic criteria, under its translation from neural terms. This is, therefore, important: if Churchland can effect a reconception, in neural terms, of epistemic virtue, then the PDP model will accommodate full-blooded (i.e. intentional) eliminativism. Unfortunately, we are told that: 'that will be no small task, and I cannot pretend seriously to undertake it here'. [78]

What Churchland *does* go on to do is 'to illustrate how some aspects of the problem can be addressed' (op cit.). Both 'explanation', and the possibility of some explanation being 'the best' explanation seem beyond accounting for, however, given the constraints under which he is now working. Thus, for example, it is observed that:

> it may be a just criticism to say that *A* is simply the *wrong* prototype for the problematic situation at hand ... because the situation confronted is not a member of the class of situations that will reliably activate *A* ... [79]

What makes an activation a 'reliable' activation is not clear: on pain of reintroducing just those semantic notions of epistemic virtue which he has undertaken to avoid, Churchland cannot suggest that, for example, favourable epistemic status accords to

some prototype activation on account of its being reliably linked to the truth. Bearing in mind that explanations are prototype activations, it's not clear what other possible criterion for 'reliability' might be proffered. Churchland is on easier ground when he chooses a relatively simple pragmatic example: adverse behavioural consequences. The hungry coyote who mistakes the tail of a poisonous snake for a retreating desert rat, and acts accordingly, dies as a result of the error. But here we have no obvious explanation (the coyote isn't explaining anything, but is rather (mis-) categorising the visual input: this is a case of faulty recognition of a 'degraded input'.[80] Where there is no explanation, there can be no consideration of whether this is the 'best' explanation - and we surely cannot take the 'reliability' criterion to be open to interpretation as 'saves one from being killed'.[81] In short, the 'coyote' example is completely irrelevant to the issue at hand.

The question of what strength of eliminative materialism is sanctioned by PDP models, as understood by Churchland, is a difficult question to answer. The mine-detector, and other artificial parallel machines, clearly do not employ propositional attitudes, nor surrogates for these - so that sentential eliminativism is true of these machines. This may or may not entail intentional eliminativism for these machines: the hidden-level activity, where the organisation of vector space into prototypical regions which will subsequently output, from any given input, in accordance with the processing of input vectors via the various prototypes, is, it seems, too far removed from sentential manipulation to be capable of being described in linguistic terms - so that intentional eliminativism is true at least at that level of the system's functioning. Whether intentional eliminativism is true of the entire system isn't clear: while the machine can

generate an output which is comprehensible to humans utilising language (it could, for example, output the text 'Warning: Mine' on a screen), it could be said that this fails to capture the full complexity of what the machine has detected: that, though it conveys all that *we* wish to know (and all that the machine is intended to achieve), it has, in deploying a prototype vector downstream of this output, identified sonar features of the object which are beyond linguistic description - so that we must not read from our parochial interests to the linguistically-available simplicity of the machine, and thus assume that intentional eliminativism is false regarding the machine.

Whether or not intentional eliminativism is true of mine-detection equipment is, however, of no great interest: what is at issue is whether it is true of human cognition - and here the issue is much more complex. As presented by Churchland, PDP processing is consistent with intentional eliminativism for both artificial and human networks. If we assume for the sake of argument that intentional eliminativism is true of the artificial networks, then whether it is also true of the human network depends upon the analogy which Churchland seeks to draw between the two. This in turn seems to depend upon the claim that the brain could support PDP processing - that nothing which is known about neurology rules out such a neural architecture; that the model is more 'psychologically realistic' - in that it will account for the speed and functional persistence of human cognition in a way that alternative models will not; and on a claimed analogy between the operation of the artificial networks, and the cognitive achievement of explanatory understanding. Neither the neurological nor the 'psychologically realistic' claims suffice to demonstrate that the brain *is* a PDP network: if true, they are merely consistent with that conclusion. On the third point -

the central claim that recognition and explanatory understanding are instances of the same cognitive achievement - this claim seems straightforwardly false, for the reasons that I have given. Genuine understanding, together with human freedom and an account of the relationship between the brain and the individual whose brain it is seem beyond the model, at least as it is currently understood by Churchland. This does not entail that no future development in PDP science might not close these explanatory gaps - though in the case of the relationship between the self and the brain, no such gap appears to be recognised by Churchland. Where difficulties are recognised - for example, the problem of prototype evaluation, which must infect any account informed by naturalistic epistemology - Churchland's position is that of Wittgenstein: 'what we cannot speak about, we must pass over in silence'.[82] There is a crucial ambiguity here: is our inability to speak about how such processes are effected in the human PDP network a principled inability (i.e. is this due to the truth of the intentional eliminativist claim that sentential description of mental states is perhaps so misleading as to be practically impossible for philosophical or scientific purposes), or is it on the grounds of absence of sufficient neuroscientific progress - in which case mere sentential eliminativism may be true? My suggestion is that the need to pass over in silence derives from the sheer empirical impossibility of the accommodation of features such as normativity, freedom, and the self, in a model which consists entirely of highly complex patterns of chemical activity. This impossibility will be reinforced in my next chapter, where I consider the prospects for a PDP account of morality.

[1] For present purposes, I propose to restrict myself to connectionism as understood by Churchland. Connectionism has appeal for many philosophers who are not eliminative materialists - and, are, indeed hostile to the eliminative materialist project. This suggests a possible range of positions on the question of what are the commitments of connectionism: attempting to take account of all of these seems neither possible nor necessary where the question in hand is what are the implications for eliminative materialism of the connectionist project, as understood by eliminativists.

[2] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', in his *A Neurocomputational Perspective - The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Mass., 1989, p.197.

[3] op. cit., p.198.

[4] I am using the term 'model' in two distinct ways here: the 'models' to which reference is made are artificial networks, whose architecture is designed along the lines postulated by the PDP model of cognition (abstractly viewed as a putative account of cognition).

[5] Similar textual idiosyncrasies have emerged elsewhere in my reading of Churchland: it may be that my pursuing these idiosyncrasies is needlessly pedantic, and unlikely to be genuinely revealing. My proposed solution is that I take Churchland's meaning literally, but endeavour not to base the conclusion of any argument on what may be mere lapses on his part.

[6] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.198.

[7] This is not, of course, to deny that there may be degrees of complexity of types to which the same token may be allocated in recognition. The biologist's categorisation of a particular individual will, for example, employ a significantly more complex taxonomy than will the layman's. Nonetheless, for any given type, a token will either be recognised or it will not.

[8] In chapter 5 I will pursue a particular instance which Churchland has raised: 'moral understanding'. The claim that moral action can be accounted for by the recognition of token instances as being members of some moral type seems to rest on two dubious premises: that understanding is 'essentially the same type of computational achievement' as recognition; and that moral choice can be reduced to some form of understanding.

[9] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.199.

[10] Even if my reading of Hempel is erroneous, only the truth of sentential eliminativism is immediately in prospect for Churchland here. The conclusion that Hempel had falsely postulated a language of thought would not entail that what *is* taking place in the mind/brain cannot be appropriately described in terms of deductive processes which quantify over propositions.

[11] It remains an open question whether it is *also* an abstract diagram of the neuronal organisation of the brain: where Churchland discusses neurobiology (as opposed to PDP research), he presents prima facie evidence of neural architecture which is consistent with the PDP model. As earlier noted, clear evidence of architectural isomorphism in the two disciplines would render Churchland's position virtually secure. It isn't clear, however, what neurobiological evidence *could* demonstrate anything stronger than consistency with the PDP model (i.e. a neural

architecture which could operate in the way postulated by PDP - but which doesn't necessarily do so).

[12] This 'configuration of synaptic weights' is a particular theory (hence my observation in the last footnote that theories are, for Churchland, neuropsychological entities rather than abstract entities). The various sets of weights which could transmit a given input vector thus represent different theories - and will ultimately eventuate in different outputs. For example, the sensory input vector generated by the sight of an elderly woman with a wart on her nose might yield the conclusion that she is a witch (the output activity may so categorise her) if the downstream synaptic weights constitute a witch-theory. The recognition that there are no witches - that this theory ought to be eliminated - will result in a reconfiguration of weights, so that a more anodyne categorisation ultimately results. In either case, the input level vector is the same - this being a function purely of the impinging of the external world on the retina. This, I take it, would constitute Churchland's account of the theory-ladenness of perception.

[13] I will cast doubt on the tenability of this account of representation below.

[14] There is, of course, the further problem that facial recognition doesn't seem like the kind of cognitive process which will utilise anything like sentential entities - and this is borne out by the divergence between the ease with which we can recognise a familiar face, and the relative difficulty in providing a verbal description of the face. If there were quasi-linguistic entities in the brain which were manipulated in cognition, then their retrieval and translation into natural language might be expected to be possible. Note how the PDP model also accommodates - as earlier noted - 'graceful degradation', and recognition from partial or distorted input data. Both these feats seem beyond the serial model, where removal of any relevant data seems likely to result in a crash of the system.

[15] It is difficult to see what a 'theory' thus is for Churchland - other than in this rather obscure sense of being a set of synaptic weights which will eventuate, given antecedent input from further downstream, in a prototype which is proprietary to that theory. It seems that both sentential and intentional eliminativism must be true of this mine-detection theory: not only is it arrived at via the manipulation of elements which are intrinsically non-sentential; the 'theory' is in principle incapable of translation into sentences (i.e. it cannot be rendered into a set of sentences which set out to describe the distinctions between the sonar echoes of rocks and of mines). Having said this, the *output* presumably *will* be capable of this translation (there would hardly, after all, be any point in having a mine detector which produced an output which couldn't be construed as meaning either 'that's a mine' or 'that's a rock'). So the hidden-level activity - where putative theory-formation takes place during the training period - is both sententially and intentionally eliminativist; the output activity is only sententially eliminativist (we could not, with any plausibility, suggest that 'that's a mine' fails to capture the full complexity of the machine's output).

[16] Intentional eliminativism is required for the already-mentioned reason that mere sentential eliminativism is endorsed by substance dualists; the need for an exhaustive account of all human cognitive faculties arises from the fact that mentalistic explanation is to be wholly eliminated by the intentional eliminativist. Churchland's allusions to such human cognitive feats as moral discrimination make clear his

appreciation of this fact - but, I will argue, his attempt to conflate simple sensory recognition and 'moral explanation' cannot be accepted.

[17] The fact that Churchland writes of the system 'recognising' or 'detecting' can't, of course, be taken to imply his commitment to mere sentential eliminativism: it is consistent with intentional eliminativism that one foresee a future mode of communication which is fundamentally different from our current natural language, but - until this is developed, one must make do with the crude and misleading verbal resources which are all that we currently have.

[18] I will, in the next chapter, raise a further difficulty in taking the mine-detector model as being analogous to such cases as human moral judgement: the 'training up' process in the mine detector is by conscious manipulation of the system's connection weights by human beings. What would be the comparable case for human moral agents? The human has, for the training set, the advantage over the mine detector, of *knowing* that *(x)* is a mine, and *(y)* is a rock - hence his ability to recognise the need for adjustment of weights. The detector's training is thus supervised from the privileged vantage point of a human mind which *knows* the difference which the detector has yet to be trained to recognise. If all human minds were similarly in need of some training up, comparable to that required by the mine detector, then it is hard to see how we could have secure basis for even this relatively simple task of recognition and differentiation. This problem bears some resemblance to the homunculus problem.

[19] In his essay 'What is a Theory of Mental Representation?', in Stephen Stich and Ted A. Warfield (eds.): *Mental Representation - A Reader*, Blackwell, Oxford, 1994.

[20] op. cit., p.352.

[21] op. cit.

[22] op. cit., p.249.

[23] op. cit. It should be noted that Stich presents this as 'early' work by Rosch - and that 'more recent research has made it clear that for many concepts ... [the prototype account] will [not] explain the data comfortably'. This need not be fatal to Churchland's use of the prototype model, however, despite his heavy reliance on vindication by current empirical research findings in the sciences, for as Stich goes on to note (p.249): 'for some concepts it has been proposed that subjects' judgements rely on something very much like a tacitly known scientific theory'. As will become clear, this possibility is congenial to Churchland's account. Churchland has elsewhere argued (in 'The Continuity of Philosophy and the Sciences' - 'Mind and Language' vol. 1, no. 1, Spring 1986) that the task for the philosopher is to be a proto-scientist, speculating about the future discoveries of science. (Consistent with this, Churchland is at pains to reject the notion that philosophy is conceptual analysis.)

[24] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach' p.207.

[25] op. cit.

[26] op. cit.

[27] op. cit., p.208.

[28] op. cit.

[29] op. cit.

[30] Given that the 'vector space' in which prototypes arise is both multidimensional and *abstract*, the claims that real brains have many more than one hidden level, and that real brains divide into distinct processing hierarchies may be hard

to substantiate. May it not be the case that these conclusions are consistent with our currently very limited neurobiological knowledge, rather than being a secure conclusion from it? Surely such phenomena as abstract multidimensional spaces may be no more than presumptive inferences from what is empirically accessible - so that there are degrees of assurance that they exist (some phenomena in this category, such as black holes, being relatively highly-assured; others, such as wormholes connecting a number of distinct universes, being highly speculative possibilities, which are no more than consistent with what is currently known).

[31] In his essay 'Explanatory Pluralism and the Co-evolution of Theories in Science', in Robert McCauley (ed.), *The Churchlands and their Critics*, Blackwell, Oxford, 1996, McCauley claims that in a 1990 jointly-authored paper, the Churchlands 'discuss and largely defuse five well-worn objections ... to the reduction of psychology to neurobiology'. One of these objections is given as 'freedom'. (op. cit. p.17). Rather disappointingly, in the paper in question ('Intertheoretic Reduction: a Neuroscientist's Field Guide', in Richard Warner and Tadeusz Szubka (eds.), *The Mind-Body Problem - A Guide to the Current Debate*, Blackwell, Oxford, 1994), the Churchlands' sole reference to the issue is to claim that: 'whether and in what sense there is any human freedom, beyond the relative autonomy that attaches to any complex dynamical system that is partially isolated from the world, is an entirely empirical question. Accordingly, rather than struggle to show that a completed neuroscience will be consistent with this, or that, or the other preconceived notion of human freedom, we recommend that we let scientific investigation *teach us* in what ways and to what degrees human creatures are "free"'. I can't agree with McCauley that this brief treatment amounts in any sense to a 'defusing' of the problem of explaining how human freedom can be accommodated in a pattern of electrochemical activity across a population of neurons. Given the claim that the philosopher is a 'proto-scientist', it seems rather a dereliction of duty to counsel that we await the future teaching of scientific investigation - especially given Churchland's bold claim that the 'stimulus-responsish' 'fear' is 'oversimple and deeply misleading' ('Explanation: A PDP Approach' p.207). I return to the issue of human freedom in the context of moral action in the next chapter.

[32] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Mass., 1995, p.3.

[33] Churchland calculates later in the same chapter (op. cit., p5) that with 'the total number of synaptic weights that the brain might assume is very roughly ten raised to the 100 trillionth power. Compare this with the measure of only 10^{87} cubic metres standardly estimated for the volume of the entire astronomical universe'. The brain thus 'encompasses a space of conceptual and cognitive possibilities that is larger ... than the entire astronomical universe' (op. cit., p.4). This rather mind-boggling statistic merely clouds the issue: does this yield the conclusion that the brain's enormity renders prediction technically impossible, or is there a source or sources of behaviour which are beyond the predictive reach of any device consistent with the laws of nature?

[34] We need also to scrutinise Churchland's assurances that the currently-private will remain so. Again the question will be whether this is due to a principled first person/ third person asymmetry, or because of insurmountable technological impediments arising from the sheer complexity of the mind/ brain, and the likely

prospects for neuroscience. I suspect that it cannot be the former - and that, by extension, the impossibility of a behaviour-predicting device is also mere technical impossibility. I return to this issue in chapter 6.

[35] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, p.101.

[36] op. cit., p.110.

[37] op. cit., p.114.

[38] Webster's Third New International Dictionary, Merriam Webster, Chicago, 1986.

[39] It isn't clear what evidence there is for this: earlier writing seems to suggest that the evidence for human cognition being via PDP is in part via analogy from artificial systems: now this is being embellished in order to attempt to accommodate specifically higher-order (e.g. human) capacities.

[40] In fact, as I will argue, more than this is required: the activation of the "danger" prototype must be *understood* by the woman.

[41] Churchland doesn't address the question of the relationship between the brain and its owner - presumably for fear of appearing to present a dualistic account. But a situation whereby the various areas of the brain have meaning only for each other (there being nothing over and above the brain) surely also fails to dispel the 'stimulus-responsish' concern (even if such internal 'meanings were possible). Who is in charge? I return to this question in later chapters.

[42] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.210.

[43] op. cit.

[44] W.V.O. Quine, 'Epistemology Naturalized' in Hilary Kornblith (ed.): *Naturalizing Epistemology*, MIT Press, Cambridge, Mass., 1994, p.25.

[45] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.210. Note, incidentally, how the term 'spontaneous' reappears here. On this occasion, it appears to mean 'using its own internal resources' - so that the connection between Quine's 'meagre input' and 'torrential output' are achieved by the network *unaided*. This seems to reinforce my earlier objection that Churchland's only tenable PDP objection to the 'stimulus-response' criticism is that there is complex internal neural activity intermediate between the stimulus and the response - not that the response is in any sense at least in part the outcome of a contributory free choice.

[46] Churchland devotes a chapter of *The Engine of the Soul, the Seat of Reason* to the question 'Could an Electronic Machine be Conscious?' - so I may be doing artificial networks some disservice in denying them the possibility of explanatory understanding. I return to the question of machine consciousness in a later chapter; for now I propose to proceed on the assumption that *extant* artificial PDP systems such as the mine detector have no capacity for explanatory understanding, and that this is to be taken to be attributable to the simplicity of their architecture.

[47] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.212.

[48] op. cit.

[49] op. cit..

[50] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.213.

[51] op. cit.

[52] op. cit. Churchland doesn't develop this point: it may be a merely infelicitous use of words on his part.

[53] op. cit.
[54] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, p.114.
[55] I will consider the Kuhnian influence on Churchland's account of the theories in the next chapter, when I come to consider Churchland's moral theory.
[56] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, p.278.
[57] op. cit., p.115.
[58] op. cit.
[59] op. cit.
[60] I assume for present purposes that the only distinction between 'explanation' and 'prediction' is temporal - explanation coming after the event, and prediction before. I also take it that Churchland would raise no objection to this assumption.
[61] Though my view, as already presented, is that he fails: his putative instances of explanation are - assuming for the moment the accuracy of the PDP account of cognition - actually instances of mere recognition/ classification.
[62] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, p.118.
[63] op. cit., p.119.
[64] paraphrased from Paul M. Churchland, 'Evaluating our Self-Conception' in *Mind and Language*, vol..8, no.2.
[65] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, p.121.
[66] I return in my final chapter to questions of truth and falsity, and the extent to which these notions might be operable within an eliminative materialist model; the question of evaluation is considered in the next chapter, where I consider the implications of Churchland's moral realism.
[67] Paul M. Churchland, 'On the Nature of Explanation: A PDP Approach', p.218.
[68] op. cit., p.219.
[69] op. cit., p.218.
[70] Consistent with his account, Churchland will eschew consideration of 'hypotheses', which are linguistic entities. In place of 'hypotheses', he will consider 'activation vectors', which perform a neural role corresponding to the logical role which hypotheses play in non-psychological accounts. Unfortunately, no such handy neural surrogate term is available for 'explanation' - so that the translation of the problem into neural terms is only partial. Churchland notes that 'in the end, the process is not one of "inference" at all, nor is its outcome generally a sentence' (op. cit.). Herein lies the problem: notwithstanding its 'crudity' from the perspective of cognitive neurodynamics, an account in which inference to that hypothesis which is maximally consistent with the facts of observation, and which thereby presents us with an 'explanation', is transparent, in that there is no mystery about what *makes* this explanation an explanation. When Churchland comes to exemplify the process from a cognitive neurodynamics perspective, the example which he gives - of a coyote being bitten by a snake - appears not to incorporate any explanation *at all* (see below).
[71] op. cit. p. 219.
[72] op. cit.
[73] In a later essay - Learning and Conceptual Change', Churchland notes that 'the (hyper)distance between the old and new prototype vectors is a measure of how great the conceptual change effected' (Paul M. Churchland, 'Learning and Conceptual Change, in his *A Neurocomputational Perspective - The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Mass., 1989, p.241).

[74] I take it that what specific population of neurons subserves a given activation is arbitrary, in much the same way that what region of a computer's hard disk contains the software which runs the computer's clock is arbitrary. This is not to ignore the fact that specific regions of the brain are allocated to specific broad types of cognitive activity (olfactory; visual; etc.). Presumably it is not held to be the case that *my* 'deflecting force' vector is activated in a location in my brain exactly equivalent to the location of the vector in Newton's and Churchland's?

[75] My impression is that Churchland's 'proximity' is proximity in something like this sense. What he seems to need, however, is this being a function of proximity in the literal - regions of the brain which are near each other - sense, as this will provide a causal account. He refers to 'response properties' of neurons as providing the 'constituting dimensions of ... very high-dimensional similarity'. Here the properties of the second of my three levels (the brain level) is accorded some obscure basis for similarity at the first level (the abstract space level) - though this is hopelessly obscure. This is both a defence ('it will typically transcend effective verbal description'), and a pressing requirement of cognitive neurodynamics: if this can be disambiguated, then it is consistent with his anti-sententialist leanings, whereas the 'semantic' (i.e. third) account leaves him once again mired in the problem of sustaining a plausible version of eliminativism - i.e. one which doesn't merely eliminate 'sentences in the head'. (All quotations in this footnote taken from 'Explanation: A PDP Approach' p.220.)

[76] op. cit., p.219.

[77] op. cit., p.220.

[78] op. cit.

[79] op. cit.

[80] Clearly we must be careful to distinguish here between the cognitive achievement of the coyote, and what *we* are capable of vis-à-vis the coyote: it isn't explaining anything in pouncing on the snake - but *we* can explain why it so ill-advisedly acted in this way. Similarly, if, per impossibile, we could communicate with the dead coyote, then it could explain its actions retrospectively - but again this doesn't entail any explanation in its brain at the point of determining to act. Once again we have the conflation of perceptual recognition and explanatory understanding: inference to the best explanation is not inference to the best feat of perceptual recognition: contra Churchland, these are not 'instances of the same form of cognitive achievement' (op. cit., p.228).

[81] Though, as I will argue in the next chapter, Churchland seems to present a not dissimilar interpretation of what makes for moral behaviour, where he cites considerations of survival and social success.

[82] Ludwig Wittgenstein, Tractatus Logico-Philosophicus, D.F. Pears and B.F. McGuinness (eds.), Routledge, London, 1961, 6.57.

CHAPTER 5 CHURCHLAND'S MORAL REALISM

Churchland's *The Engine of Reason, the Seat of the Soul* [1] demonstrates the full Kuhnian scope of his ambition. The project has shifted from being a (possibly) radical contribution to philosophy of mind, to being a putative revolution in human thought generally. This synoptic ambition underlies the text, whose aim is:

> to make available ... the character and potential significance of the developing theory and the recent experimental results ... [the reader] will be better able to participate in the inevitable debates about appropriate public policy concerning medical care, psychiatry, the law, moral responsibility, our correctional system, education, private morality, and the nature of freedom ... it is ... crucial that relevant information be made widely available.[2]

Churchland's synoptic ambition for PDP-based eliminative materialism provides the motivation for his recent work on moral philosophy. The claims that the brain is nothing but a PDP system, and the physicalist claim that the mind is the brain, jointly entail that all of our mental lives can be accounted for in terms of activation vectors. I will refer to this as the 'radical synoptic claim': the claim that *all* human mental activity is the processing of a PDP brain. Churchland must thus argue that moral judgement is an example of mental activity which can be accommodated within the PDP model. He must show how a moral position is consistent with eliminativist PDP.

One of the ways of taxonomising the standard moral positions is to distinguish moral realism, moral irrealism, and moral nihilism. Of these three, irrealism seems most clearly impossible to reconcile with PDP. In summarising the irrealist position to which emotivists, prescriptivists and projectivists will all subscribe, Smith notes that the

irrealist will deny that there are moral facts, insisting rather that 'our moral judgements simply express our desires about how people behave'.[3] For the irrealist, 'desire' is central to morality. It is unclear what Churchland's position on desire is: his attack on folk psychology is exclusively in terms of the other half of what is sometimes referred to as 'belief-desire psychology'. No configuration of the variables of the PDP model (synaptic weights; activation vectors; the partitioning of neuronal activation space) seem likely to be obvious candidates for some neuronal surrogate for desire.[4] In my last chapter I drew attention to Churchland's rejection of the will (at least as traditionally conceived); the related traditional element of desire would appear to be among what Smith terms 'queer moral properties', which include will, and which are intrinsic to moral irrealism.[5]

If Smith's taxonomy of moral positions is exhaustive (as it appears to be), then Churchland is thus left with two options: either he can endorse moral nihilism, or he can attempt to accommodate moral realism in his PDP model. It is unclear why Churchland doesn't avail himself of the moral nihilist option, which is consistent with a particularly radical form of eliminative materialism. Given that our traditional self-conception is put in question by Churchland's thesis, it could be argued that part of that self-conception is of persons as moral agents - but that, as moral agency seems ineluctably tied to the discredited ontology of folk psychology, the alternative conception which the neurocomputational perspective entails is so radical as to discard even moral agency. It may be the case that Churchland finds this intuitively unappealing (though intuition is itself a rather conventional basis for philosophical conviction). As with the 'free will' issue, it appears to be Churchland's evident wish to

maximise the appeal of the putative paradigm shift in our self-conception which leads him to attempt to avoid the moral nihilist option.[6] The claim that morality is a sham - and the consequent implication that, for example, the holocaust cannot be appropriately described as having been a moral wrong - will be unpersuasive to the intended lay audience for his *The Engine of Reason, the Seat of the Soul*.

Much thus depends on Churchland's ability to sustain a PDP version of moral realism. If he fails, then his position must embrace the massively counter-intuitive moral nihilism. The only other alternative would be to accept that no PDP account of morality is possible - but that this does not impugn the status of morality. But for Churchland to adopt this option is for him to concede defeat: he needs PDP to entail eliminativism. If the 'queer properties' of moral irrealism are bona fide properties which can be neither accommodated by PDP (i.e. reduced) nor eliminated, then eliminative materialism is not entailed by what empirical success PDP may have in other domains. In the absence of any alternative to PDP which will successfully reduce the 'queer properties', explanation in terms of properties and states which are essentially non-physical is neither reducible nor eliminable - so that the central claim of eliminative materialism is false.[7] The question is thus whether there can be a PDP account of morality - that is, a PDP version of moral realism - and what consequences the account will have for the further question of what strength of commitment to eliminative materialism Churchland can endorse.

Churchland's rejection of traditional epistemology precludes his endorsement of a conventional form of moral realism, whereby one discovers putative moral truths via

the exercise of reason, which uncovers moral laws[8]. Such an aprioristic route to moral truth must be replaced by one which is empiricist - and consistent with the PDP model, given the synoptic claim regarding this model. Thus moral knowledge, like scientific knowledge, must be embodied. Churchland's moral concepts will be prototypes. This recasting of moral realism is consistent with what I will term 'the continuity thesis'. Churchland's claim will be that the only difference between moral knowledge and scientific knowledge is that moral prototypes are activated in the deployment of the former, but not in the latter.

Churchland's first published work on moral philosophy is in his *A Neurocomputational Perspective*, which ends with a short chapter entitled 'Moral Facts and Moral Knowledge': 'moral truths, I shall argue, are roughly as robust and objective as other instances of truth ...'.[9] In order to evaluate the case for Churchland's putative PDP-based morality - the case for 'moral knowledge' being coded in patterns of vectorial activity in the brain - it will be helpful to consider first, what Churchland takes 'knowledge' to consist in, and secondly, what he takes 'morality' to consist in.

Clearly, the epistemology that makes eliminative materialism possible must deviate from traditional accounts of knowledge as 'justified true belief'. The claim that these three conditions are individually necessary and jointly sufficient for knowledge assumes a traditional notion of concepts as having precisely specifiable necessary and sufficient conditions: as noted earlier, the empirical work of psychologists such as Rosch, which forms part of the context for the development of the notion of PDP

prototypes, postulates concepts which are not amenable to such precise definition. Like the sonar echoes of mines, concepts will have loose, family-resemblance relations between their individual instantiations. But there are further motivations for Churchland to reject the 'justified true belief' construal: most conspicuous is the fact that knowledge must on the PDP account be *embodied*, and non-sentential. For a cognitive component - in this case, a neural component - to satisfy the condition of being *justified* and *true* would appear ineluctably for it to be a component which enters into just the kind of semantic relations which Churchland must eschew. The 'justified true belief' condition seems to be consistent with, at most, sentential eliminativism - there are no sentences in the head, but the chemical activity which there *is* eventuates in states which are translatable into sentences which can then be evaluated for justifiability and truth. Churchland cannot without difficulty accept the notion of 'truth' in his account of knowledge:

> although the history of human intellectual endeavour does support the view that over the centuries our theories have become dramatically *better* in many dimensions, it is quite problematic whether they are successfully "closer" to "truth". Indeed, the notion of truth itself has recently come in for critical scrutiny ... it is no longer clear that there *is* any unique and unitary relation that virtuous belief systems must bear to the nonlinguistic world.[10]

What we have here is *at least* the bracketing of 'truth', on the grounds that there may be a number of possible criteria for epistemic virtue.[11]

Conventional accounts of knowledge - those consistent with the 'justified true belief' condition - will account for an individual's having knowledge in terms of that individual's knowing that ... (where what will conventionally replace the ellipsis is a

proposition - so that the conventional account seems ineluctably linked to propositional attitudes). By contrast, Churchland emphasises 'knowing-how', in alluding to:

> ... the practical or pragmatic nature of both scientific and broadly normative knowledge ... both embody different forms of *know-how*: how to navigate the natural world in the former case, and how to navigate the social world in the latter.[12]

So knowledge is a brain state, the possession of which facilitates a certain performance on the part of the individual whose brain state it is, and the evaluation of which must be in non-traditional (i.e. non-semantic) terms. To have 'knowledge' in a particular domain is to have a brain which is configured with just those synaptic weights which will trigger prototypes relevant to the performance in question.

This takes us some way towards understanding 'the epistemology that makes eliminative materialism possible'. Churchland acknowledges an intellectual debt to Kuhn - and his reading of Kuhn's *The Structure of Scientific Revolutions* is presented by Churchland in defence of his epistemological position. Thus, for example, in his *The Engine of Reason, the Seat of the Soul*, Churchland states that Kuhn's text 'upset my own Logical Empiricist assumptions':

> it had that effect for two reasons. The first was his claim ... that past scientific revolutions were not the unambiguous expression of sheerly logical and experimental factors ... [but] were the expression of ... social, psychological, metaphysical, technological, aesthetic, and personal [factors] ... the second reason ... was his claim ... that the unit of scientific understanding is not the sentence, or set of sentences, but the so-called "paradigm", or family of paradigms.[13]

The latter of these Kuhnian contributions may be quickly dispensed with for my present purposes. Churchland sees the Kuhnian 'paradigm' as being a prototype: in both cases, what we have is a scientific revolution eventuating in the novel deployment of this paradigm/ prototype - this constituting an explanatory achievement. In consequence, we have a new development in 'know-how'. Clearly, if the unit of scientific understanding is not the sentence, then we have sentential eliminativism for theories (at least). The former of Kuhn's claims is more important for the purposes of the present chapter (where what is at issue is the possibility of embodied, moral, facts). Churchland's argument for the continuity thesis (the thesis that there is no fundamental distinction between science and morals) will be based largely on putative analogies between scientific progress and moral progress - so that the criteria for evaluation (social, psychological, etc.) which are here outlined will be shown to be operative in the case of moral development of the community. Thus, for example, Kuhn:

> urged a "performance" conception of theory evaluation ... a theory is a vehicle whose virtue lies in its many uses ... the evaluation of a theory by the scientific community is almost always a matter of complex social and intellectual negotiation.[14]

Given this negotiability, deriving from the claim that 'it is no longer clear that there *is* any unique and unitary relation that virtuous belief systems must bear to the nonlinguistic world', it is unclear why Churchland alludes to 'moral *facts*'. 'Facts' are surely nonnegotiable: once established, they may be open to various novel *interpretations* - but the facts *per se* will remain unchanged. More seriously, a fact would appear to be an entity which is open to sentential description: it can be 'picked

out' by means of a sentence of the form: 'it is a fact that *a planet's curved path in three-dimensional space is in reality a straight path within the non-Euclidean geometry of the four-dimensional spacetime continuum that surrounds the "attracting body"*.[15] If the items of 'knowledge' are to be immune to semantic interpretation, and if they are to be elements in a syndrome which is open to continuing challenge via continuing intellectual progress, then these items are surely inappropriately dubbed 'facts'.

Having briefly considered the first of the two questions raised earlier (what does Churchland take 'knowledge' to be?), I now propose to go on to the second question: what does Churchland take 'morality' to consist in?

For the continuity thesis to be sustained, the only distinction between 'scientific knowledge' and 'moral knowledge' which can be accepted is the distinction that the former is via prototypes which are not moral prototypes. Thus, given that a Kuhnian, performance-evaluable construal of scientific knowledge is advocated, so 'moral knowledge' must also be 'know-how': as noted earlier, moral knowledge can be defined as: 'how to navigate the social world'.[16] As Churchland notes, this account of morality contrasts sharply with:

> ... the more traditional accounts that picture the moral person as one who has agreed to follow a certain set of *rules* (e.g., "Always keep your promises", etc.), or alternatively, as one who has a certain set of overriding *desires* (e.g., to maximise the general happiness, etc.). Both of these more traditional accounts are badly out of focus.[17]

The final point here may be taken to indicate the linguistic challenge posed by Churchland's account: were he to claim that the traditional accounts were *false*, then this would appear to entail the presumed *truth* of his own account - hence the need to resort to metaphor ('badly out of focus'), in order to obviate assessment in terms of an epistemology other than that which makes eliminative materialism possible.

In reinforcing the first part of this claim - the claim that morality does not consist in the internalisation, and subsequent following of, rules, Churchland moves from sentential to intentional eliminativism:

> ... it is just not possible to capture, in a set of explicit imperative sentences or rules, more than a small part of the practical wisdom possessed by a mature moral individual ... statable rules are not the *basis* of one's moral character. They are merely its pale and partial reflection at the comparatively impotent level of language.[18]

Churchland is concerned, not merely to deny that the moral agent is possessed of a set of internal 'sentences in the head' such as 'Always keep your promises', but that, in addition, to claim that what *is* in the head - the syndrome of synaptic weights and prototype vectors which constitutes the individual's moral 'expertise' - can be reflected in only a 'pale and partial' manner by the 'impotent' resources of language. This entails intentional eliminativism: strictly speaking, sentential descriptions of internal cognitive states ought not to be utilised.

The second part of the claim in Churchland's quotation - that the notion of the moral person as one who acts on desires is 'badly out of focus' - leads to the claim that:

> a person might have an all-consuming desire to maximise human happiness. But if that person has no comprehension of what sorts of things genuinely serve lasting human happiness; no capacity for recognising other people's emotions, aspirations and current purposes ... no skills whatever at pursuing that all-consuming desire; then that person is not a moral saint. He is a pathetic fool ...[19]

The opening observation here, that the person in question 'might have ... desire' of the requisite sort, is ambiguous: it could mean either that it is possible for someone to be in this condition; or, alternatively, it may not mean this at all - and be intended merely as granting the precondition which would have to obtain in order for the opposing point of view to work, prior to demonstrating that it does not, in fact, work as an account of morality. Given my earlier observations regarding the absence of any conspicuous neural surrogate for desire in Churchland's PDP model, I take it that the latter is the appropriate interpretation - and that Churchland is not conceding the possibility of someone's being in the folk psychological-sounding condition of 'overriding desire'. But setting aside this quibble, what follows in the quotation is revealing of Churchland's conception of 'morality'. Consistent with his performative conception of knowledge - and hence of moral knowledge, Churchland's position is that the *intention* of the agent is of no relevance in the evaluation of his actions; what *is* relevant is the agent's knowledge of how to promote happiness within the community. The corollary of this is that the shrewd and calculating individual, who manipulates public opinion in his own interest by pandering to the public's perceived interests, along the lines of the sophists and orators in Plato's dialogues, will, if successful in their assessment of the public mood, be candidates for 'moral sainthood'. [20] To be morally good is to be an effective performer, consistent with moral

knowledge being knowing how to act morally. The corollary of this is that the inept performer is 'a fool'.

Having denied that desire to act morally is sufficient for morality, Churchland goes on to deny its necessity:

> a man may have, as his most basic and overriding desire in life, the desire to see his own children mature and prosper. To him, let us suppose, everything else is distantly secondary. And yet, such a person may still be numbered among the most consummately moral people of his community, as long as he pursues his personal goal ... in a fashion that is scrupulously fair to the aspirations of others ...[21]

This seems a rather weak position, given the clear importance for Churchland's account of his presenting criteria for the evaluation of moral performance which will fulfil the same function as the Kuhnian criteria for scientific knowledge. 'Scrupulous fairness to the aspirations of others' cannot, in any case, be taken as a universalisable criterion: the criteria for successful performance must be indexed to the specific context within which the individual is operating. Churchland's moral realism thus differs from most traditional accounts in a further respect: its relativism. There can be no moral knowledge which is not indexed to the prevailing local conditions of the individual's community. Churchland responds to the question "why be moral?":

> as well ask, "Why should I acquire the skills of swimming?" when one is a fish. In both cases, the short answer is, "Consider, dear creature, the environment in which you have no choice but to live." To be sure, this answer leaves open the question of exactly what motor skills will make one the *best possible* swimmer, and likewise the question of exactly what social skills will make one the *maximally successful* social agent.[22]

Once again, Churchland's position leaves no room for what one intends by one's actions: maximal social success is the sole performance criterion. But the '[social] environment in which you have no choice but to live' may well establish criteria for maximal social success which are, contra Churchland's rather blasé formulation, conspicuously hostile to the objective of being 'scrupulously fair to the aspirations of others' . Pastor Niemoller, a Berlin churchman in the 1930s, founded a group which helped combat rising discrimination against Christians of Jewish background under the Nazis. This led to Niemoller's arrest by the Gestapo, and eventual incarceration in the Dachau concentration camp. How will Churchland evaluate this behaviour? Niemoller famously (and rhetorically) observed that:

> when Hitler attacked the Jews I was not a Jew, therefore I was not concerned. And when Hitler attacked the Catholics, I was not a Catholic, and therefore, I was not concerned. And when Hitler attacked the unions and the industrialists, I was not a member of the unions and I was not concerned. Then, Hitler attacked me and the Protestant church - and there was nobody left to be concerned.[23]

Clearly, when the majority - perhaps the entirety - of the community is opposed to one's publicly-adopted position, then one's actions are conspicuously not conducive to one's being 'maximally socially successful'. Churchland thus has two options: either he can brand Niemoller 'a pathetic fool' for his adoption of a position so radically at odds with 'the emotions, aspirations, current purposes ...' of the community - the social environment - in which he lived; alternatively, he must widen the scope of this community, both geographically and temporally, beyond 1930s Germany - so that Niemoller's actions *are* consonant with the human community in the modern age, or some similar formulation of 'community'.[24] This now poses a problem, however:

unless we can delineate 'the moral environment' in such a way as to redeem Niemoller - but without encompassing the conflict of purposes and aspirations etc. which a plurality of cultures and time periods will likely generate - then it will become impossible both to accommodate the undoubtedly moral actions of Niemoller, *and* have a coherent set of community-based criteria by which to index an individual's morality or immorality.[25] The only other alternative would be to adopt rules for morality which are timeless and not culturally-specific - and Churchland has already discounted such a basis for morality.

Thus far, I have presented only Churchland's desire for explanatory unity as a motivation for this reconception of morality as knowledge embodied in the brain, and the consequent 'social skill' conception of morality. There are, however, a number of analogies which Churchland sees between the scientific and the moral domains. As analogy plays a significant role in his PDP epistemology - it is the basis for scientific progress, as problematic cases are incorporated under novel prototypes, hitherto employed for the analogous case - he can be expected to find such analogies persuasive.[26]

Churchland identifies three respects in which moral knowledge is analogous to scientific knowledge: the state of the layman's knowledge in each case; the achievement of progress in the two domains; and the criteria for evaluating knowledge in the two domains.

The first of these analogies - that concerned with the state of the layman's knowledge - arises in the course of Churchland's response to an anticipated objection to the continuity claim: that scientific theory possesses a history of progress, and an inherent objectivity, which stands in sharp contrast to the 'flimsy, arbitrary and subjective' moral and social domain.[27] Churchland's response to this is to claim that a false analogy is being posed: the appropriate analogue for the layman's confused, narrow and arbitrary moral and social convictions, is, not the 'carefully distilled wisdom of institutionalised science', but the equally confused, narrow and arbitrary scientific knowledge of the layman. Churchland concludes:

> if anything, the average person displays a slightly higher level of moral cognition than of scientific cognition. If we wish to disqualify moral cognition as a form of knowledge, we must look to some other contrasts to bring it down.[28]

It's difficult to see how Churchland believes that anything is demonstrated by this observation - let alone that it supports the continuity thesis. Even if it were admissible to internalise theories and knowledge so that they have no abstract existence - in fact, no existence at all - outside human brains, it would still be inadmissible to deduce anything about the objective nature of these entities from the layman's limited grasp of them. The argument then begs the question by assuming that people's 'ignorance, prejudice, self-interest, class interest, unbridled emotion and religious enthusiasm' are evidence of a low level of moral *knowledge.*[29] His response misses the point, in any case. What makes possible the advancement of science towards the 'carefully distilled wisdom' of the professional scientist is precisely the fact that there is what Churchland elsewhere terms 'an objective configuration of objects and properties' to which true

scientific claims will correspond.[30] The claim that knowledge consists in knowing how to do something - and that morality and science are, despite claims to the contrary, not fundamentally different, *on the grounds that we don't do either particularly well*, cannot seriously be held to support the case for the continuity thesis.

As already noted, a key element in Churchland's PDP model of explanation is the account which he offers of scientific progress. Churchland sees a comparable process of continuous adjustment of social policy as further evidence of the close analogy between the moral and scientific domains. The recurrent nature of individual human cognition is recalled in the process of the accumulation of legislation via 'long experience and many adjustments'.[31] Just as there is no infallibility in science - even Newton's prototypes were superseded by the better explanation advanced by Einstein, whose theory is itself now a candidate for similar displacement - so there is an ongoing need for adjustment in social policy. The basis for evaluation is - in keeping with Churchland's Kuhnian epistemology - wholly pragmatic: whether a given policy aids our collective operation in the social environment. Again, however, Churchland's analogy here seems not to sustain the continuity thesis in the way that he intends. In the case of continuous adjustments of social policy, what we have is a set of laws and institutional arrangements, rather than a body of moral knowledge. The prevailing policy will have moral implications, of course - but continuous adjustment of social policy is not the same thing as continuous adjustment of moral convictions or moral knowledge. The analogy becomes even more strained when Churchland introduces the judiciary - whose role in interpretation of the abstract wisdom of the legislature is likened by him to that of engineers, whose task is to apply abstract wisdom to actual

cases. The analogue of the scientific paradigm - the prototype - is in this case the judicial precedent. Lawyers and judges thus attempt to apprehend a great diversity of actual cases 'as instances of some antecedent prototype'.[32] The merit of this arrangement, as Churchland points out, is that it ensures consistency in the application of the law.

At this point, a disanalogy between this process, and the research efforts of scientists and engineers, arguably becomes apparent. Consistency is an objective in law because only by being consistent can the law be *fair* - consistent treatment of all who are subject to the law is one of the central principles of the rule of law, which is a feature of a fair society. So there is an intrinsically normative dimension to the task facing legislators and judges in a liberal democratic society: consistent treatment is fair - therefore it is *good.* In the domain of science, 'consistency' is regarded as virtuous - but here there is no objective of being 'fair'. Consistent handling of control variables, and consistent treatment of raw research data ensures the reliability of the ultimate findings, and is thus a constraint on good scientific procedure (where 'good' means 'effective').[33] I conclude that, to the extent that this analogy succeeds, it does so on account of the specifically *non-moral* (and hence non-normative) aspects of the process of legislation and legal interpretation.

The gap between morals and the law is conceded by Churchland, when he observes that:

> the realm of socially-enforced law encodes only the most serious of our collective convictions about appropriate and inappropriate behaviour. Beneath that realm there is a similar body of shared understanding, a similar framework of social-recognitional and social-behavioural skills that we expect our fellow citizens to command.[34]

Given the social dislocation of contemporary Western society - nowhere more apparent than in the United States - this seems an extraordinary claim. Neither the apparent suggestion that there is general consensus regarding the 'most serious of our ... convictions', nor the claim that there is further consensus at a lower level, stand up to any scrutiny. This is important, given the central claim of the continuity thesis. As my 'Pastor Neimoller' example was intended to demonstrate, Churchland's 'maximally successful social agent' is one who can act in consonance with the objectives and priorities of the wider community - so that there must be a fact of the matter about what this communal set of objectives and priorities *is*. The analogy here is with the transition from one paradigm to another in Kuhnian theory: here the transition 'is a social fact accomplished by a community of scientists'.[35] There is thus a pressing need for the claim which Churchland goes on to make apropos the law and judicial precedent: 'this is the domain of common public morality'.[36] In the absence of any such common morality, there is little prospect of the individual's being able to learn how to operate as a maximally successful social agent, where this is indexed to criteria putatively established at this communal level. The absence of the accomplishment of such a communal set of prototypes will also weaken the analogy with science, as the latter is conceived by Kuhn and Churchland.[37]

In my previous chapter I raised Churchland's own formulation of an objection to his PDP model of explanation, to the effect that the model will accommodate only recognition and subsequent classification, rather than explanation.[38] In the case of moral explanation, a further problem is consequent upon Churchland's continuity thesis. In virtue of what is a process of classification and explanation a process of *moral* classification and explanation? This problem is evident in an example which Churchland cites in defence of his model - that of abortion:

> one side of the debate considers the status of the early foetus and invokes the moral prototype of a Person, albeit a very tiny and incomplete person ... the other side of the debate addresses the same situation and invokes the prototype of a tiny and possibly unwelcome Growth, as yet no more a person than is a cyst or a cluster of one's own skin cells.[39]

As it stands, this is a clear example of recognition. As with the judge's activation of a 'judicial precedent' prototype, there is no explanation here (reinforcing my claim that not all prototype activations are explanations, even if all explanations are prototype activations). In the case of the activation of a scientific prototype, there is, I suggest, a gap for which Churchland fails to account, between the activation of the prototype, and the subsequent explanation which is given as an output. In the moral case, the gap will be between classification - the activation of one or other of the prototypes 'Person' or 'Unwanted Growth' - and action. Churchland's account of the woman leaving the burning building will be in terms of output vectors being in this case motor vectors - so that muscular activity is stimulated as output, and she runs away. In the 'foetus' case, the output could be either a judgement: 'abortion is permissible/ not permissible in this case', or it could be an action - the action of seeking an abortion, or

continuing with the pregnancy.[40] The problem is in saying what makes the present case a moral case *at all.* The issue of the status of the foetus is surely a metaphysical, rather than a moral issue: if the foetus is categorised as a 'person' - then, given certain possible construals of what is a person, moral status may derive from this metaphysical status.[41] If we consider the case of a gynaecologist's prototype activation on meeting a woman who requests an abortion, and assume that the gynaecologist's 'Unwanted Growth' prototype is activated, then motor activity will be stimulated in what I take it will be a comparable manner to that in the case of the woman's running from the building *(mutatis mutandis).* In that case, Churchland's vaunted analogy between moral cognition and scientific cognition has been so successful, that it is no longer clear from this account what makes a judgement a *moral* judgement. If a candidate in a medical science examination is shown a sample of tissue in a jar of preserving fluid, and asked to identify the contents of the jar, and if the candidate determines that the object in the jar is a growth of some kind - say, a cyst - then the PDP account will allude to the activation of a 'Growth' prototype in the candidate's brain, leading to a motor output, as the candidate lifts his arm to the answer paper, and writes 'growth' in the space provided for this sample. Churchland would appear to have no account of what makes the *gynaecologist's* judgement a *moral* judgement, and the *medical student's* judgement a *scientific* judgement. If this interpretation is accurate, then Churchland has effectively eliminated morality by so successfully conflating the moral domain with that of science (given the presumed accuracy of the PDP account) that the boundary between the two disappears on his account.[42]

What is missing from Churchland's account is the fact that the metaphysical question of the status of the foetus has moral force in virtue of the fact that its answer entails the further, and moral, judgement that 'if I have an unwanted foetus, then I ought/ ought not to terminate the pregnancy'.[43] The difference between morals and science is precisely this deployment of the moral 'ought': the claim that morals and science differ only in terms of the nature of the prototype activated will only stand if this difference is somehow accommodated: in the absence of normativity, there can apparently be no difference, so that morality is effectively eliminated. Nor can we say that the moral nature of the gynaecologist's prototype can be read off from the moral nature of her subsequent *action*. While this might seem prima facie consistent with the performative conception of moral knowledge postulated by Churchland, such a manoeuvre merely serves to shift the problem, as we must now ask what, on his account, makes an *action* a *moral* action? We cannot claim that the determining factor is the judgement of third parties, whose prototype firings on apprehending her action are moral prototype firings, as this merely returns us to the question of what is a moral prototype. 'Social' success - conformity with the actions of the rest of the community - seems similarly too ambiguous to serve as an effective criterion for the demarcation of the moral action from the non-moral, and the consequent identification of the antecedent prototype as a moral prototype. Ants, for example, behave in a manner which is consistent with a maximally-successful ant community, with elaborate divisions of labour and hierarchies, and a sophisticated network of interdependence. I take it that no one will claim that the effective soldier ant's behaviour is therefore morally good - so that 'maximal social success' cannot, without further clarification, serve to delineate the moral from the non-moral.

Roger Scruton raises a question which he takes to be 'the crux for the moral realist':

> what is to be said to someone who agrees with the moral argument, but feels no inclination to act on it?[44]

The attempt to construct a possible PDP response to this question raises interesting consequences for Churchland's moral realism, given his concern to resist the objection that his model cannot accommodate human freedom. Scruton's formulation appears conventionally sentential: 'agreement with an argument' is a semantic condition, and assumes knowledge-that, rather than know-how. In place of such agreement, the PDP theorist will, I take it, characterise the situation as one of comparable prototypes (or the same prototype) being activated in both brains, or all of the brains in question. The question of whether the same prototype might instigate different outputs in different brains is an interesting one. Churchland's account gives no reason to expect that the same prototype *will* always eventuate in the same type of output response - and this would appear to be confirmable a posteriori, by the empirical fact that, for example, a pervert's response to a given prototype activation in his brain may be exactly the opposite of the response of 'normal' individuals.[45] Scruton's question appears to be based on the rationally-characterisable situation where one may correctly *deduce*, from the combination of some moral rule and antecedent circumstances, how one ought to act. Aristotle's notion of 'incontinence' - where I know how I ought to act, having worked out the appropriate practical syllogism, *but elect not to do so*, would appear to capture the kind of case which Scruton has in mind. But the 'queer moral properties' of the moral irrealist again come into play here. 'Inclination' is a condition which

doesn't appear capable of being transferred to the PDP model: I can, on the Scruton and Aristotelian traditional models, *will myself* to act or not to act in accordance with some judgement regarding how I ought to act or ought not to act. But once again, Churchland's account appears 'stimulus-responsish': once all of the appropriate prototypes have been activated at the hidden levels in my brain, my behavioural output is rendered inevitable by the synaptic connections and synaptic connection strengths already established in the brain in the abstract region of these prototypes. Interestingly, Churchland appears to have a response to Scruton: Aristotelian incontinence cannot arise, so that the objection to moral realism is misplaced. The downstream vector activations are associated with the appropriate motor activity. This is, however, a Pyrrhic victory, as Churchland has in consequence no obvious room in his model for 'inclinations to act'.[46] It is thus completely unclear how Churchland could respond to the question which Scruton sets for the realist.

The earlier-noted question: 'what is a person?' is of interest in the light of this evident gap in Churchland's account. The tradition, which he has so conspicuously bracketed on various occasions, will postulate the possession of reason as a necessary condition. [47] Churchland's account of morality must therefore be an account which can accommodate morality in the absence of rationality.[48] This raises the questions: could a PDP machine be constructed which made moral choices; and if not, why not? Suppose we were to select some society where there was consensus at all levels (government, church, individual) that the foetus is a person - and emphatically *not* a mere growth. A machine very like the mine detector could then be trained up to recognise foetuses and growths - as a result of an input of x-ray images from the

womb. The internal connection weights in the network are initially set at random values, and the training set of recorded images - half of them foetuses, and half cysts - are fed into the machine's input layer. Using the backpropagation technique of synaptic weight adjustment, we cycle repeatedly through the training set until the machine outputs the message 'don't operate' in response to each of the foetuses in the training set, and 'operate' in response to each of the cysts. The training set is enormous, and includes many non-standard cysts (and foetuses); its ultimate performance is comparable to the mine-detector in terms of successful identification of inputs.

What we now have is a machine which is maximally efficient, in that its outputs conform precisely to the objectives, priorities, and preferences of the community. In terms of Churchland's performance criteria for moral knowledge, the machine is 'consummately successful'. It is unclear to me how Churchland can reject the claim that the machine is making moral judgements - or acting morally.[49] The only distinctions between the machine's operation and human operation are (a) their respective physical compositions - which is clearly irrelevant; and (b) the complexity of their respective internal operations: the 'foetus-detector' makes use of neither recurrent nor horizontal processing: like the mine-detector, it is an exclusively 'feed-forward' network. This latter distinction cannot be held to be what makes an output a moral judgement.[50] Just as computational complexity intermediate between the input and the eventual output cannot of itself account for human freedom, nor can it account for the presence of a moral judgement: where two networks produce identical outputs from identical inputs, and in identical contexts, (and where

neither network's functioning is being augmented by an additional, non-PDP, component) the claim that the output of one has the property of being a moral judgement, while the output of the other does not, is not credible. I conclude that Churchland must concede the possibility of artificial networks making moral choices. [51]

When Churchland comes to consider the issue of moral wrongdoing, further interesting light is shed on his position. The 'moral miscreant' is an individual who is 'unskilled in social practices'.[52] The account which he offers of this moral miscreancy appeals in the first instance to lack of appropriate training - so that the prototypes for moral behaviour which is compatible with social success are poorly developed. The miscreant in this case presumably resembles a mine-detector which is still undergoing the training-up which will eventually lead to its successful performance, but whose performance is currently deficient. But the question then arises of what one ought to do about this regrettable situation. In a chapter in *The Engine of Reason* entitled 'The Brain in Trouble', Churchland considers how therapy might operate, in accordance with the PDP model. In particular, he contrasts 'talk' and 'chemical and surgical intervention'.[53]

Churchland's account of the perceived failures of Freudian analysis are presented in terms of his general model of competition between theories. Freud is credited with a feat of scientific creativity, in that he 'attempted to redeploy the central family of *commonsense* cognitive prototypes - beliefs, desires, fears, and practical reasoning - in a new domain: the Unconscious'.[54] Freudianism is thus a paradigmatic example of

the novel deployment of already-familiar prototypes. Given that these are the prototypes of folk psychology, Churchland is, not surprisingly, critical of the theory, and can cite poor performance - the standard Kuhnian criterion - in defence of his position. Churchland's analysis of the source of this theoretical failure appeals to sentential eliminativism: cognitive processes lack the sentence-like and inference-like structure of the commonsense prototypes for belief, desire, fear, and practical reasoning. As already urged, activation vectors operate via vectorial transformation - so that the pathological behaviour which is the target of therapy has not arisen as a result of sequences of inferential stress defined over propositions. Churchland concludes that:

> ... we have done far better by looking for structural failures or abnormalities in the brain, for functional failures in its physiology, for chemical abnormalities in its metabolism, for genetic failures in its original blueprint, and for developmental hitches in its maturation.[55]

This claim must have enormous consequences for the prospects of Churchland being able to offer a plausible account of either human freedom (a response to the 'stimulus-responsish' criticism), or the human capacity for moral choice. Churchland's conclusion is that: '[we cannot] fix a genuinely broken brain just by talking to it'.[56]

Given the claim just made - that functional, chemical, and structural anomalies in the brain are the likely sources of pathological behaviour, and the vectorial and transformative nature of the brain's functioning (whether defective or not), then this latter claim must be true, where a 'broken brain' is one which generates sub-optimal social practices as its output. Given that Churchland has raised the dichotomy between

'talk' and 'chemical and surgical intervention' in this section of the text, it must surely be the case that chemical or surgical - reconstructive - intervention will be the most causally effective procedure, where rectification of the problem - the adaptation of behaviour to conform with appropriate social practices - is the objective.

Churchland hesitates to endorse this conclusion, however. He observes that 'drugs or surgery might enable the process, but only social interaction can actually provide [the solution]'.[57] This, however, must be held to be inconsistent: *if* it is the case that all moral reasoning consists solely of vector-to-vector transformation in the brain, where the causal processes subserving this vectorial activity are reducible to chemical causal sequences which have the end result that they do in virtue of the global configuration of connections which obtain within the brain, then surely a therapy which consisted *solely* of either adjustment of the brain's chemical balance, or adjustment of the synaptic weights, would be effective in adjusting the behavioural output.[58] 'Talk' might make a contribution to the process, but it wouldn't be strictly necessary.[59] It may be objected, of course, that human processing is much more complex - being massively recurrent and horizontal as well as 'feed forward' in its operation. Once again, though, human cognitive complexity would be called upon to rescue eliminative materialism, having already been postulated by me as a possible eliminativist response to the 'stimulus-responsish' objection, and the moral machine objection.[60]

The clear implication of this is, of course, the return of the 'stimulus-responsish' problem in an apparently intractable form: by introducing a chemical into the public water supply, I could presumably effect a change in brain state which could in turn

influence the behaviour of those who ingest the chemical. If the behavioural change took the form of a shift from sub-optimal - to - optimal social practice, then this would have to be construed on the Churchland account as an enhancement in the moral status of the individuals concerned. This is a preposterous assumption - but, as with the assumption that a machine could be constructed which could make moral choices, it isn't clear how Churchland can reject it while retaining a position consistent with his PDP model.

My claim is that - within the terms of reference of the PDP model - chemical intervention would have to be deemed *more effective* than 'talk'. Churchland suggests that 'there will always be a place for systematic conversation ... in the therapeutic process'. But this linguistic process will have to be received as input by the system - in its auditory input region, as a set of acoustic phenomena - then processed via vectorial transformation to the relevant prototypes. This is a deeply mysterious process, as it is still unclear *how* the brain can effect the conversion of vibrations in the air, via chemical transformations in the brain, to activate subsequent upstream vectorial activity which has the appropriate meaning (i.e. which represents the same thing(s) as the representations in the brain of the therapist - is the *same prototype*). There is both enormous scope for cognitive dissonance - a slightly different prototype being activated in the therapist's and the patient's respective brains (this being especially likely given the nature of the patient's perceived problem in the first place) - and the likely need for a very large amount of this painstaking hit-and-miss activation, in order to nudge the brain into developing the pathways which will ensure socially successful behavioural outputs.[61]

My suspicion is that Churchland is here failing to follow his own argument to its inevitable conclusion - that what the future may hold is a world of moral readjustment by pharmaceutical intervention - because the claim is likely to be too radical to have appeal to his audience (the layman for whom his *The Engine of Reason, the Seat of the Soul* is principally intended).[62]

Churchland's room for manoeuvre here is limited. It is difficult to see how adjustment of the brain's chemistry could *fail* to have the desired effect on the behavioural outputs of the human PDP system, in the absence of any other 'therapy'. On the other hand, standards for moral behaviour must be indexed to a particular community, on Churchland's account.[63] The achievement of moral conformity via purely pharmaceutical means now seems implausible - as the drug would require to be adjusted to the specific moral standards of the community in question - so that the emphasis on 'talk' rescues the account from absurdity. Churchland thus has a dilemma: if he is to remain consistent with his claim that brain chemistry is the basis for moral choice, then he needs to concede that drug therapy would be sufficient. But for drug therapy to be sufficient, the set of behaviours would require to be more standardised than his moral-relativism model assumes. The possibility of one pill for adjustment to Muslim fundamentalist morality; another for Ulster loyalism; and so on, seems too ridiculous to be given any consideration. But Churchland must retain the moral relativism model, in order to be able to identify Kuhnian performance criteria. There patently are no universal standards of actual moral practice - so, in the absence of rationally-discoverable moral rules, from which actual moral practice may, or may not,

deviate (and he has rejected this possibility on Kuhnian epistemological grounds), then there must be the relativity of 'moral facts'. Given the absurdity of the different-pills-for-different-local-moral-standards notion, he is constrained to fall back on the therapeutic virtues of 'talk' - notwithstanding his withering assessment of Freud, for whom therapy consisted *exclusively* in talk.[64]

Once this is conceded, the prospects for any form of eliminative materialism stronger than mere sentential eliminativism seem poor. Moral education - or moral re-education - via 'talk' will take the form of sentential inputs, with patterns of inference, and, in all likelihood, the normative terminology of the rejected epistemology. The reconfiguration of 'talk' so as to eschew these features seems impossible (so that the miscreant will be appealed to on grounds such as "if you wish to be maximally successful socially, then you should behave in accordance with standards of behaviour expected by the community; surely you want to be socially successful because ..."). If this is the input, then what results from this in terms of hidden- and output-level transformations, would appear to be capable of sentential description with a fair degree of descriptive accuracy, contra the intentional eliminativist claim, even if it is true that the processing involved is intrinsically non-sentential.

At the beginning of this chapter, I set out what I have termed the 'radical synoptic claim' - that all human cognitive activity, including the formulation of moral judgement, may be accounted for in purely PDP terms. If this claim is true, then future scientific research will enhance the neurocomputational perspective until all aspects of morals, social policy, and aesthetics, as well as science, will be accounted for purely in

terms of the operation of human PDP networks. The claim that all cognitive activity is PDP activity entails the weaker claim that human sensory recognition is by PDP processing. The claim that the PDP account of human brain functioning is *totally* false will thus include the claim that sensory recognition is not subserved by vectorial activation in the brain. I do not wish to make this claim, as it seems prima facie plausible to claim that procedures such as facial recognition are achieved, at least in part, by a process which is replicated by the artificial mine detector.[65] But the claim that human sensory recognition is subserved by a PDP process in the brain is consistent with substance dualism (the self acting as the counterpart of the sonar operator who uses the mine detector) - so that my willingness to accept the possible accuracy of this claim does not entail acceptance of an outcome remotely near the radical synoptic claim advanced in Churchland's *The Engine of Reason, the Seat of the Soul*. If PDP is to entail eliminative materialism, then the radical synoptic claim must be true.

It remains to be proven that moral choice *could* be exercised by an exclusively PDP network. Churchland has left implicit throughout the apparent assumption that the only distinction between scientific and moral 'explanation' is the nature of the prototypes whose activation is that explanation. But the key question of what makes a prototype a moral prototype is left unaddressed, the apparent assumption being that one can read off the moral nature of the prototype in question from the categorisation of the action to which it gives rise. This could work in one of two ways: either we could deduce the activation of a moral prototype from the individual's attempt to operate in a 'socially successful' manner; alternatively, we could accept that the action is, by common

consent, action within a moral domain (as in the case of abortion). The first of these options fails to provide a pragmatic answer to the question of what makes a prototype a moral prototype, in virtue of the fact that 'social success' is too ambiguous a category, incorporating as it does both human behaviours which are morally neutral, and animal behaviours. The latter option is circular: what is accepted by common consent to be moral action will be determined by the prototypes which are commonly deployed in apprehending such action; we would still require an account of what makes *these* prototypes 'moral'.

What is thus at issue is Churchland's ability to sustain his claim that his position is *moral* realism. The key element which I suggest is missing is a normative element: if Churchland can accept that the difference between a moral prototype and a non-moral prototype is that the former, but not the latter, explain how we *ought* to behave, then he can answer the question of what makes a prototype a moral prototype. If, however, he cannot accommodate this distinction between the moral and the non-moral - and if the ineliminability of the moral 'ought' is accepted - then morality itself ought to be eliminated, so that, despite his professed endorsement of moral realism, Churchland's position is, in fact, moral nihilist.

Moral nihilism is consistent with intentional eliminativism, so that in endorsing moral nihilism, Churchland would keep open the option of a genuinely eliminative materialism. The claim that moral judgement is an example of mental activity which it is impossible to accommodate within the PDP model, so that the synoptic claim for PDP is false, would be addressed by the counterclaim that there is no domain of moral

judgement, so that the success already achieved in duplicating other forms of human cognition in artificial PDP systems secures the possibility of an exclusively PDP account of all human cognition, this account availing itself of none of the explanatory categories of folk psychology. But his desire to render his general position acceptable to the widest possible audience presents a dilemma, for the converse is also true: if Churchland endorses normativity as an essential feature of moral judgement, then only sentential eliminativism seems to be in prospect. If we describe some individual's moral stance as being a commitment to the rightness or wrongness of a particular practice, such as abortion, then we have normativity - but it is difficult to see how this sentential presentation could be fundamentally wrong - or even wrong in any significant respect. In this case, Churchland will have gained a satisfactory account of what makes a judgement a moral judgement - but at the cost of relinquishing all but the modest sentential form of eliminativism.

[1] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul - a Philosophical Journey into the Brain*, MIT Press, Cambridge, Mass., 1995

[2] Churchland, *The Engine of Reason*, p.19.

[3] Michael Smith, 'Realism', in Peter Singer (ed.): *A Companion to Ethics*, Blackwell, London, 1991, p.402.

[4] The same claim may be made for belief. Churchland tends to consider belief in terms of his opposition to propositional attitude psychology - so that anti-sententialism (and hence at least sentential eliminativism) is being advocated. It isn't clear whether Churchland wishes to eliminate knowledge-that - or whether it is merely to be bracketed pending future PDP developments. In either case, the knowing-how alternative would appear incapable of accommodating belief, which must be belief-that, and hence must stand in a close epistemic relationship to knowledge-that.

[5] Smith, 'Realism' p.404.

[6] This is an unorthodox ambition for a scientific revolutionary - and seems not to have been a consideration for Copernicus or Newton, for example.

[7] It would, of course, remain possible that there *is* some non-PDP account of how the brain works, which will entail eliminative materialism, and which accommodates moral realism. But Churchland's commitment to PDP is such that this would be at best an extremely tenuous speculative support for eliminative materialism.

[8] Thus, while post-Quinean epistemology does not entail moral realism per se, one who subscribes to post-Quineanism and who is in addition a moral realist cannot endorse the discovery of putative moral truths via reason. I make this point as Churchland is himself concerned to distance his position from such rationalist-based moral realism, but nonetheless to locate himself within moral realism. There is, of course, a British tradition of eighteenth century moral-sense realists - such as Shaftesbury, Hutcheson and Hume, who occupy just this position of moral realism which has an empiricist basis.

[9] Paul M. Churchland, 'Moral Facts and Moral Knowledge', in his *A Neurocomputational Perspective - The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Massachusetts, 1989, p.297.

[10] Paul M. Churchland, 'On the Nature of Theories', in his *A Neurocomputational Perspective*, p.157.

[11] Once again the sentential eliminativist/ intentional eliminativist dichotomy comes into play here: 'truth', as conventionally understood, is a relationship which a proposition bears - whether to the nonlinguistic world, as in correspondence accounts, or to other propositions, as in coherence accounts. Acceptance of either of these accounts would entail the acceptance of, at most, mere sentential eliminativism.

[12] Paul M. Churchland, *The Engine of Reason*, p.292.

[13] op. cit., p.272.

[14] op. cit., p.276. Churchland cites an interesting example in his 'Learning and Conceptual Change': classical thermodynamics was enormously successful - the evidence of this being that 'it helped to produce the industrial revolution' (in *A Neurocomputational Perspective*, p.242). So we have a performance criterion for theory evaluation. But this success didn't prevent Bernoulli et al from attempting to reconceive thermal phenomena under the prototypes germane to kinematic and

corpuscular theories. Churchland doesn't allude here to progress towards *truth* - the implication seems to be that the criterion for success in the case of classical thermodynamics (aiding the industrial revolution) differs from the criterion for success in subsequent theory (theoretical unification) - so that the criteria for successful performance are not fixed - but are, presumably, negotiable in terms of the criteria - social, psychological, etc. - set out by Churchland.

[15] Example taken from Churchland: *The Engine of Reason*, p.119 (Einstein's 'still better' interpretation of the phenomena of planetary behaviour, as previously interpreted by Newton).

[16] op. cit., p.292.

[17] op. cit., p.293.

[18] op. cit., p.293.

[19] op. cit., p.293.

[20] Churchland does rather seem to miss the point of morality in excluding all consideration of motivation - but this is necessitated by his anti-irrealism, which is in turn the consequence of the need to eschew 'queer properties'. In any case, a similar objection could be raised against consequentialists who are not eliminative materialists, and my task here is not to criticise consequentialism per se.

[21] Paul M. Churchland, *The Engine of Reason*, p.293.

[22] op. cit., p.150.

[23] quoted in A. Partington (ed.), *The Oxford Dictionary of Quotations*, Revised 4th edition, Oxford University Press, Oxford, 1996, p.495.

[24] Despite my profound scepticism regarding his philosophical positions, I don't doubt that Churchland is a humane and decent man - who would, I am sure, have as much respect for Niemoller's self-sacrifice as do I. This narrows Churchland's options to one - the second here cited.

[25] It would seem that any such delineation will be arbitrary - and certainly one which was contrived in such a way as to square the circle which I here identify would lack persuasiveness, as it would appear to be motivated by nothing more than explanatory convenience. The question of what comprises a 'moral environment' or 'social environment' is also problematic in the opposite direction: Nazi Germany in the 1930s was unusual for the uniformity of the moral climate which was consequent upon the imposition of totalitarian rule by the Nazis. In contemporary pluralist society, there would appear to be no obvious consensus about what are the objectives, emotions, priorities etc. of 'the community'. Indeed, a recent British prime minister famously observed that 'there is no such thing as society' - so that the social aggregation of purposes and priorities postulated by Churchland as a prelude to establishing what will constitute morally-skilled behaviour - and hence, ipso facto, moral knowledge - may not be available at all.

[26] Of course, given his epistemological convictions, Churchland will not be concerned by the observation that '*(a)* is analogous to *(b)*' does not entail that '*(a)* is the same as *(b)* in all relevant respects'; nor will he be overly concerned by objections to the effect that an analogy is 'not close': as suggested in my previous chapter, it is not entirely clear what constitutes analogical 'closeness' on the PDP account - but in any case, the genius of thinkers such as Aristotle and Newton lay precisely in their grasping the possibility of the consideration of a problematic case under prototypes which were 'distant' from those conventionally deployed. In short, the claim that an

analogy does not amount to evidence will require some criterion for evidence which doesn't beg the question against Churchland by assuming traditional epistemology.

[27] Paul M. Churchland, *The Engine of Reason*, p.287.

[28] op. cit.

[29] op. cit.

[30] Paul M. Churchland, *A Neurocomputational Perspective*, p.297.

[31] Paul M. Churchland, *The Engine of Reason*, p.288.

[32] op. cit., p.290.

[33] Churchland's allusion here to 'consistency' is, I take it, intended to capture the fact that, once a prototype is established - and until its replacement, whether by act of parliament or new judicial precedent - all broadly similar problematic cases are apprehended as instances of the same prototype, in the same way that, in the periods between Kuhnian 'paradigm shifts', scientists will similarly apprehend all broadly similar events or phenomena under the same prototype. This is a fair comparison - but where the intention is to emphasise analogy between the two, it is important to recognise that this very criterion of 'consistency' is the basis for very significant divergence between the domains of law and science. Churchland's legal example also serves to reinforce a point which I raised in my last chapter (p) - where I pointed out that while all explanatory understanding consists in prototype activation on the model, this does not entail that all prototype activation is explanatory understanding (some room must be made for mere *classification*, and this would *also* have to be prototype activation). On this basis, it seemed unfair to parody the ancients' dubbing of the constellations as 'the dipper'; 'the great bear', etc., as putative *explanations* of the underlying phenomena. Similarly, the activation in the judge's brain of a given prototype, under which he will apprehend a given case, does not amount to that judge's *explanation* of the case - but rather his classification of the type of case which the case in question is.

[34] Paul M Churchland, *The Engine of Reason*, p.290.

[35] Wesley C. Salmon, '[The] Epistemology of Natural Science' in Jonathan Dancy and Ernest Sosa (eds.): *A Companion to Epistemology*, Blackwell, Oxford, 1992, p.295.

[36] Paul M. Churchland, *The Engine of Reason*, p.291.

[37] Churchland appears here to be trying to have his epistemological cake and eat it: an anarchic diversity of moral prototypes being employed across the community, with little if any moral consensus, would be contrary to the existence of 'moral facts' - to which he alludes in the title of the closing chapter ('Moral Facts and Moral Knowledge') of his *A Neurocomputational Perspective*. If we adhere to a conventional epistemology - so that to have moral knowledge consists in having justified true belief that we ought to, for example, 'Always keep our promises' - then the lack of communal agreement would lack justification: those who took issue with this rule would be *wrong*. But the epistemology upon which Churchland's eliminative materialism is based ought - as previously noted - to eschew 'moral facts', as being targets of 'knowledge-that', and hence sententialist; as being apparently unrevisable (contrary to the commitment to indefinite progress in our knowledge); and as being true (truth having been bracketed pending further progress in PDP epistemology).

[38] p.109. One is inclined to claim that in the context of morals, the objection seems even stronger: the process of prototype activation in response to the

apprehension of a particular case or event - seems again capable only of accounting for recognition of a particular moral type, but not of accounting for what must surely be the ultimate point of moral cognition - *moral action*. My immediate task is not, however, to criticise moral realism per se - but rather its putative PDP formulation.

[39] Paul M. Churchland, *The Engine of Reason*, p.147.

[40] There may be further complications here: how does the PDP model account for the distinction between those moral prototype activations which eventuate in judgement, and those which eventuate in action? It isn't clear whether much hangs on this distinction - but I cannot see how it will be accounted for using the resources of PDP.

[41] Given the potential vagueness of prototypes/ concepts in the PDP model, the model might be thought to gain credence here, given the notorious difficulty in specifying necessary and sufficient conditions for personhood - and I take it that this is in part Churchland's point here: it will be, in the absence of such necessary and sufficient conditions, a 'judgement call' which prototype is activated. But presumably this same lack of precision will infect the question of the moral status of a particular person - so that we cannot rely upon both vagueness and precision in arriving at the conclusion that (a) it is open to debate whether a foetus is a person, but (b) if it is, then it has the same moral status as all other persons. This seems directly analogous to Roschian psychological studies which suggest that some birds - such as robins - are unambiguously birds, while others - such as emus - are birds, but their status as such is more ambiguous.

[42] Part of the PDP moral realist model which Churchland has constructed seems in any case to be missing here. In keeping with a Kuhnian epistemology, there must be some performance criterion for evaluating the 'moral knowledge' which is being deployed in judging that abortion is permissible/ impermissible. How does one act in order to satisfy the maximally successful social agent criterion in this case? It may be, of course, that some alternative performance criterion which is consistent with the successful navigation of the social world is available (Kuhnian theory, as presented by Churchland, doesn't privilege any one criterion for evaluation over all others) - but it isn't clear what this might be. In any case, this highly ambiguous situation sits ill at ease with the claim by Churchland that there are 'moral facts'.

[43] The entailment from the metaphysical claim to the moral claim will require the insertion of a further premise(s). This could be just the kind of rule ('persons' right to life ought to be respected') or desire (e.g. the desire to maximise the number of potentially happy lives in existence) which Churchland has earlier rejected. This, I think, is the problem: his concern to account for *how moral decisions are reached*, and to render this account consistent with his PDP model, has required that he strip out just those elements which enable us to state *what morality is*. Moral judgements *are* moral judgements in part due to the fact that they provide us with reasons for acting in a particular way. The bracketing of rationality precludes this characterisation by Churchland.

[44] Roger Scruton, *Modern Philosophy - A Survey*, Sinclair-Stevenson, London, 1994, p.561.

[45] It isn't entirely clear whether this example works. In the same way in which there must be more vectorial activity in the woman's brain when she is in the burning building than activation of the 'Smoke' prototype (how else would we account for her

not fleeing the barbecue?), similarly, the pervert's prototype must either be accompanied by *further* prototypes with which they are associated in his brain - and which in turn give rise to the gratification which he experiences, *or* the prototype must be qualitatively different in the first place. The former seems more consistent with Churchland's account (i.e. further prototypes being activated in conjunction with the one in question). I consider later in this chapter Churchland's observations regarding the scope for treatment of deviant individuals, given the apparent opportunities afforded by the neurological research of which he avails himself.

[46] One is tempted to add that once again morality seems to be under threat of elimination on Churchland's account: moral actions must surely arise from one's free choice to act morally. What made Niemoller's actions morally courageous was precisely the fact that he willed himself to act against his own self-interest: had his actions been the consequence merely of chemical activity in his brain, then their moral force would have been lost. But this would be to bring to bear against Churchland's position an anti-consequentialist objection, rather than an anti- PDP moral realism objection.

[47] See, e.g., the opening section of his *The Engine of Reason*; also his revealingly-titled 'Evaluating Our Self-Conception', in *Mind & Language,* vol.8, no.2, Summer 1993. In fairness to Churchland's position of scepticism regarding the possibility of 'classical' definitions which stipulate necessary and sufficient conditions for their instantiation, it is not unproblematically clear that absence of reason disqualifies one from entitlement to consideration as a 'person'. Young babies and patients in a persistently vegetative state lack reason - but one is disinclined to claim that they are 'not persons'

[48] Andy Clark suggests that 'reflection on the form of inner encoding might ... cause us to alter or expand *some* elements of our conception of rationality ... Churchland ... suggestively argues that once we understand the non-sentential (prototype-involving) form of inner economy we will come to place knowing-how (rather than knowing-that) at the centre of our vision of reasoning and rationality.' ('The Varieties of Eliminativism' in *Mind & Language* vol.8, no.2, Summer 1993, p.229). I find this set of claims rather vague. The connection between 'knowing-that' and rationality is clear (and is exemplified in Aristotle's practical syllogism). But it isn't clear how *any* of the elements of our conception of rationality might survive the elimination of this type of knowledge, and its replacement with knowing-how. Clark goes on to suggest that 'perhaps ... we may come to place less weight on reconstructing decisions (or moral judgements) as the conclusions of sequences of inferential stress defined over propositions and to instead treat reasoning and decision-making more like *e.g.* driving a car' (op. cit., p.230). This last quotation seems to rest on an equivocation surrounding the term 'reasoning'. If reasoning means 'working out what to do' - then it is arguable that Churchland's mine detector does this, as it processes the input vectors which arise from sonar echoes (the machine determines whether to output 'Mine' or 'Rock' on its screen). This is 'knowing how' in Churchland's rather strained sense of 'knowing how to react' (though I cannot see how this differs from '*knowing that* the object is a mine'). Clark's example of 'driving a car' reinforces my point that there is a clear and fundamental distinction between 'reasoning' in this loose sense, and rational choice. What is it to be rational when driving a car? It is surely to have a set of antecedent objectives (get to where one is

intending to go; don't hit any pedestrians; don't violate any traffic laws; etc.) - and to behave in a way which is consistent with the satisfaction of these antecedent objectives. The sentential eliminativist claim - that I don't process 'sequences of inferential stress defined over propositions' in my head while driving - may well be true. But if intentional eliminativism is true - so that my driving behaviour cannot be *reconstructed or interpreted* as 'if you want to avoid killing anyone, don't drive on the pavement' - so that my keeping on the road is the consequence of my not wanting to kill anyone, together with my knowledge that if I drive on the pavement I am liable to kill someone - then rationality is surely lost.

[49] I will ignore for present purposes the usual problem of describing an intrinsically non-sentential system's operation without recourse to potentially-misleading natural language resources - also the fact that moral action is left mysterious in the case of human PDP systems.

[50] The gynaecologist and the medical student of my earlier example will have brains of comparable architectural complexity: given the similarity of their training, their prototypes can also be assumed to be at least very similar. Yet one was making a moral judgement, the other an ontological judgement in response to the input.

[51] In my next chapter I will consider the related question of whether a machine can be conscious. I will there argue that the absence of the possibility of subjective experience of a machine - the principled absence of a first-person perspective - entails the impossibility of a machine's behaviour being appropriately characterised as 'moral' behaviour. I take it that, just as Niemoller must be free for his actions to have value, for the same reason, he must be conscious for the actions to be morally evaluable.

[52] Paul M. Churchland, *The Engine of Reason*, p.149. Such a characterisation - studiously avoiding the employment of normative terms such as 'bad' or 'wrong' - serves to highlight once again the difficulty in dichotomising between the moral and the non-moral in Churchland's account. The expression 'unskilled in social practices' is redolent of the terminology employed by the finishing school. Someone who doesn't know how to eat shellfish in polite company isn't immoral - but is arguably 'unskilled in social practices'. On a less flippant note, sufferers from autism may have, as one of their conspicuous symptoms, a lack of skill in social practices. Churchland needs to be able to exempt such unfortunate people from the category of the 'moral miscreant' - but it isn't clear how this can be achieved, without recourse to the 'queer properties' of the moral irrealist ('it isn't their *fault*, they don't *choose* to behave in this way').

[53] Paul M. Churchland, *The Engine of Reason*, p.181 et. seq.

[54] op. cit., p.182.

[55] op. cit., p.183.

[56] op. cit.

[57] op. cit.

[58] By 'adjustment of synaptic weights' I mean the process carried out on the mine detector during training up. *It* doesn't have to be spoken to (indeed, speaking to it would hardly be of any benefit). So my assumption here is that - at least in principle - comparable backpropagation would be possible for human 'training up'.

[59] This issue is of enormous significance for the question of how strong a thesis Churchland's eliminative materialism is to be. If talk has a role, then this seems inevitably to weaken the thesis: given that talk may also constitute the output (e.g. on the therapist's couch), then the presence of talk as both an input and and output must

limit the extent to which we can say that internal states intermediate between these are inaccurately individuated in terms of their propositional content - so that only mere sentential eliminativism is in prospect. If on the other hand, talk can in future be dispensed with in the case of therapy, then its *general* dispensability is in prospect - and this is an assumption central to a particularly extreme form of eliminative materialism. I consider the possibility of the general elimination of 'talk' in my concluding chapter.

[60] In any case, the 'impossibility' which might be appealed to by the eliminativist in this case could surely be no more than mere technical impossibility - impossibility within the context of what is currently-available technology.

[61] I assume that this construction of new pathways is possible, however - and that the brain of the patient is here replicating approximately the same process as the creative or scientific process, where the individual thinks 'laterally', and deploys novel pathways in utilising prototype vectors which will later become second-nature.

[62] As I will discuss in my concluding chapter, there is a radical version of eliminative materialism - which Churchland has considered briefly in his earlier writing - which would involve the elimination of all talk, whether as a vehicle for psychological therapy, or for any other purpose.

[63] The Churchlands offer a suitably naturalistic explanation of this: 'as with the ecology of a tropical rainforest ... the human ecology is exquisitely complex and can take many different forms, depending on the local economic circumstances, social organisations, technological developments, geographic location, and idiosyncratic history. In the case of natural ecology, it is obvious that the behavioural "skills" that will make an individual fern a flourishing and well-integrated ecological citizen are utterly dependent on the details of that enveloping natural ecology. In this sense, the fern's "biological virtues" are always "ecosystem relative"'. ('Flanagan on Moral Knowledge' in Robert N. McCauley (ed.), *The Churchlands and their Critics*, Blackwell, Oxford, 1996, p.303).

[64] Herein lies the problem: Freud is both lampooned by Churchland (' ... the bearded and monocled Freudian analyst probing his reclining patient for memories of toilet training gone awry and parentally directed lust') (*The Engine of Reason*, p.181), and criticised, on the more serious grounds of his employing a defective (because folk-psychological) prototype as his explanation of the internal workings of the mind. But only the latter, it seems, is being rejected by Churchland: the 'talk' part of the therapy remains. The therapy presumably would consist less in asking the patient questions - whether about toilet training or parental lust, or anything else - and more in telling the patient how he ought to behave. This rather unappealing scenario (one could presumably accommodate the Gestapo's remonstrations with Niemoller in this model - they were advising him on how to be maximally successful as a social agent in Nazi society) departs only slightly from Freudian therapy - which must be a disappointment for those of a radical eliminativist disposition.

[65] My objection that the mine detector does not interpret its own output, so that an interpreter and controller of the process is still required, still stands - so that I do not accept that the mine detector research demonstrates that we are in some sense sophisticated mine detectors. The significance of what the mine detector outputs is significance to or for some operator; by extension, the achievement of a PDP

recognition in network in the human brain would be output which has significance - not to the network itself, but - to the self whose brain it is.

CHAPTER 6 CAN CHURCHLAND ACCOUNT FOR CONSCIOUSNESS?

Churchland is concerned to tackle an argument which is advanced, in different ways, by Nagel and Jackson, who:

> argue for the existence of a special, intrinsically perspectival kind of *fact*, the fact of "what it is like" to have a mental experience of such-and-such a kind, which intractably and in principle cannot be captured or explained by physical science.[1]

Nagel proposes that this fact - of there being something that it is like to be a particular organism - is linked to that organism's being conscious:

> ... the fact that an organism has conscious experience *at all* means, basically, that there is something that it is like to *be* that organism.[2]

If it is a necessary condition for being conscious that a being instantiates this condition, then the absence of the condition is sufficient for the absence of consciousness. If the condition in question is in principle intractable to physical science, as suggested above, then consciousness is ipso facto intractable to science. This entails in turn the falsity of the claim that all mentalistic explanation, together with its ontology, is either reducible to physicalistic explanation, or is eliminable - so that Nagel's claims, if true, are fatal to any but the most modest forms of eliminative materialism.

In this chapter, I will first consider Churchland's argument against Nagel and Jackson, before going on to consider his account of how consciousness may be accommodated

by his PDP model - even to the extent that a conscious machine is an empirical possibility.

Churchland expresses the problem which he must address as being that of a claimed fundamental bifurcation between the objective and the subjective:

> Consciousness, it has been argued, is essentially a *subjective* phenomenon, accessible only to the creature that has it, while anything that is truly physical - one's brain activity, for example - is ... *objective* in nature, that is, to be accessible to many people from many points of view.[3]

Both Churchlands have rightly objected to what amounts to an *argumentum ad ignorantiam*, to the effect that as we cannot *at present* provide an account of consciousness using only the explanatory resources of physical science, no such account *can* be given. Thus Patricia Churchland observes that:

> beginning with Thomas Nagel, various philosophers have proposed setting conscious experience apart from all other problems of the mind as "the most difficult problem". When critically examined, the basis for this proposal reveals itself to be ... use of our current ignorance.[4]

The 'ignorance' in question is ignorance with respect to the question of how an objective account of 'what it is like to be' could be provided. A similar argument - based on a putative inference from what cannot be *imagined* - is attacked at some length by Paul Churchland in his *The Engine of Reason*. Characteristically, he rebuts this type of argument by appeal to a number of instances from the history of science. Thus, for example, Ptolemy apparently rejected the possibility of our ever being able to provide genuinely scientific explanation of the nature and motions of the stars and

planets, on account of their distance, and the consequent impossibility of imagining this impediment being overcome. The solution to Ptolemy's problem is presented in PDP terms:

> [what Ptolemy lacked was] ... 'the conceptual framework that Newton would later construct ... Newton's framework ... would have partitioned parts of Ptolemy's neuronal activation space in a new and radically different way.[5]

This is, of course, an anticipation of what Churchland himself will later go on to attempt in his positive account of consciousness: the radical re-partitioning of our neuronal activation space, in order to render the currently unimaginable (to some) physical science account of consciousness imaginable. In the meantime, both Churchlands do appear to be attacking straw men - for the argument for the principled intractability of our subjective mental lives to physical science is based on much more effective arguments than either current ignorance or unimaginability. There is, in any case, a countervailing logical trap, which Churchland recognises:

> to be fair, however, neither can substantive theoretical questions be decided just by citing some carefully chosen examples from the history of science, examples that may, or may not, be genuine parallels to the case at issue ... [6]

In preparation for his response to the Nagel objection, Churchland makes use of a section from Leibniz's *Monadology*:

> suppose that there were a machine so constructed as to produce thought, feeling, and perception, we could imagine it increased in size while retaining the same proportions, so that one could enter as one might a mill. On going inside we should only see the parts impinging upon one another; we should not see anything which would explain a perception.[7]

Churchland's response to this is to deny that it supports the conclusion which Leibniz intends (namely, the dualistic conclusion that 'thought, feeling and perception' belong to a different order of reality from the physical parts of the mechanism which 'produces' them): 'it remains possible, even granting Leibniz's story, that the taste sensation of a peach is identical with a four-element activation vector in the gustatory pathways'.[8] Leibniz's thought experiment is, he suggests, another example of *argumentam ad ignorantiam*: in the absence of possession of the necessary explanatory resources - that is, the appropriately partitioned neural activation space - we *would* fail to 'see anything which would explain a perception'. This again anticipates Churchland's later approach to the topic of consciousness: if he can contribute to the requisite partitioning, then his claim is that just such explanation will be possible with regard to study of the human brain: an exhaustive account of the taste of a peach will be possible from our suitably-trained examination of the relevant gustatory activation vector. Nagel's argument will be less easily dismissed, however, as it makes explicit one of the features of the taste of a peach which such an exhaustive account would require to include: *what it is like* to taste a peach. Churchland's claim that it is not impossible for suitably-trained examination of the gustatory vector to yield what Leibniz terms 'explanation' of a perception must include the claim that this property - of there being 'something that it is like' to have the experience which is the vector activation - will also be subject to exhaustive and exclusively physical explanation. Nagel's claim is that this is known to the individual whose experience it is in a subjective manner which renders it beyond any possible physical - and thus objective and publicly-accessible - account.

Churchland accepts one of the elements in Nagel's argument - namely, that there is a way of knowing which is intrinsic and unique to the knower. This is explained as:

> a unique set of intimate *causal connections* to the sensory activity of one's own brain and nervous system ... what this means is that each creature has a way of knowing about its *own* sensory states that no other creature has ... other people cannot know [your sensory states] ... via the individual informational pathways by which *you* know them, because only you possess exactly those pathways.[9]

Churchland thus recognises an epistemic asymmetry. The claim that there is this asymmetry does not of itself entail that there is something which can be known *only* via these 'auto-connected pathways' in the brain. Such an outcome would result in first-person privilege with regard to the item of knowledge in question. The presence of particulars which are uniquely proprietary to this mode of knowing might be thought further to entail their intrinsically non-physical nature, on the grounds that were they to be physical, then they would be ipso facto accessible to third-person inspection (though perhaps only given suitable neuroscientific training). While the existence of this way of knowing - via auto-connected pathways - does not entail the existence of non-physical individuals, it is nonetheless consistent with their existence, as Churchland recognises:

> Does the undoubted existence of this unique way of your knowing about your own internal states mean that there is something non-physical about those states, something that must transcend representation within physical science? Perhaps. [10]

If we take 'physical science' to be the totality of physical explanation, which is in turn the totality of physical prototypes, then the question at issue is whether, given an ideally-completed set of all possible physical prototypes, the phenomenon in question would be appropriately apprehended under one of these prototypes.

While an asymmetry in the fundamental nature of things known entails an asymmetry of modes of knowing (the presence of mental particulars which are subject to first-person privilege will entail a non-physical - and hence non-scientific - mode of knowing them); the entailment does not work in the opposite direction: an asymmetry of modes of knowing does not entail an asymmetry in the fundamental nature of things known. Were there to be such an entailment, then Churchland's recognition of the asymmetry of modes of knowing would compel his endorsement of dualism. Churchland is concerned to stress this point, via a number of cases where there is putatively just such a divergence: epistemic asymmetry, but no fundamental metaphysical asymmetry:

> by way of an axonal network called your *proprioceptive system*, you have informational access to the physical configuration of your own body and limbs. That information comes from the millions of sensors in your muscles, sensors that convey tension information to the brain. Nobody else can know the configuration of your body in this particular way. Only you. Because only your brain enjoys the relevant causal connections to your body. Others must use other means to know your bodily configuration: they must see it, or feel it with their hands ... *the object of knowledge is the same from both perspectives*, the subjective and the objective. (my emphasis)[11]

Two problems arise from Churchland's use of this account of the proprioceptive system. Firstly, the 'multitude' of similar examples which he goes on to indicate (such

as 'the fullness of your bladder, and of your bowels') (op cit.) do not collectively amount to an argument for the claim that epistemic asymmetry *never* accompanies an asymmetry in the fundamental nature of things known: 'some' does not entail 'all' - so that the observation (if true) that there is epistemic asymmetry in the absence of asymmetry in the fundamental nature of things known in this multitude of cases is too weak to serve the purpose for which it is apparently intended - that is, to demonstrate that such an accompaniment of epistemic *and* metaphysical asymmetry *never* occurs. Secondly, and more seriously, Churchland begs the question here: we need not accept the observation that the fullness of one's bladder is subject to epistemic, but not objective asymmetry ('the object of knowledge is the same from both perspectives') - and hence need not accept Churchland's further conclusion that: 'there is nothing supraphysical, nothing beyond the bounds of physical science here'.[12]

Apropos the earlier example of the configuration of one's limbs, Churchland claims that the object of knowledge in both cases (i.e. the subjective/ proprioceptive and objective/ third-person) ' ... is something paradigmatically physical ... the configuration of your body and limbs'.[13] Clearly the configuration of one's body and limbs *per se* is indeed physical. But it remains to be proved that this is '*the* object of knowledge' - that is, that the object of knowledge is *nothing but* - this physical configuration. In the same way that there is something that it is like to be a bat using echolocation, so there is something that it is like to have just my current configuration of limbs. The dualist will not accept that this may be included in the category of the 'paradigmatically physical' - so that the claim is question-begging.

All that Churchland *can* legitimately claim up to this point is that, as already noted, a divergence of ways of knowing does not *entail* a difference in what is known - but this modest claim does not refute any claim which Nagel and Jackson make. His conclusion that: 'the existence of a proprietary, first-person epistemological access to some phenomenon does not mean that the accessed phenomenon is non-physical in nature' can therefore be accepted without any damage being done to the Nagel position.

One option open to Churchland - but, somewhat surprisingly, not considered by him explicitly in his *The Engine of Reason* - is to sustain the claim that there are epistemic asymmetries but no corresponding metaphysical asymmetries on the grounds that knowledge is knowledge *how* rather than knowledge *that*. Robinson describes this claim (i.e. the claim that all knowledge is knowledge how) as a 'behaviouristic intuition'.[14] The argument might then proceed by claiming that what is at issue - the putatively non-physical entities - would be targets for knowledge that, but that in the absence of knowledge that, all knowledge is within the category of 'modes of knowledge'; there is no separate category of knowledge which would satisfy the dualist's intuition that there is something which the scientist lacks vis-à-vis his knowledge of the configuration of my limbs - knowledge that it is like *(x)* for me (where what would stand in place of *(x)* here is ineffable on account of the essentially non-public nature of the experience). It is difficult to know how to evaluate this possible line of argument. The emphasis on knowledge-how in Churchland's writing is motivated not merely by his Kuhnian influence, but also by his desire to resist sententialism. If knowledge is knowledge-that, then there would appear to be no

prospect of any stronger formulation than sentential eliminativism being sustainable: patterns of activation activity in the brain, which are intrinsically non-sentential but which generate a condition appropriately described by a propositional attitude, cannot be an attractive proposition for Churchland. Robinson's allusion to behaviourism is also useful in this context: the avoidance of the subjective which characterises behaviourism will lead to the attempt to eschew such mentalistic-sounding formulations as the individual being in a state characterisable as being 'knowledge that'. The argument that the ineluctably subjective can thus be eliminated will in consequence be question-begging, as an anti-subjectivist epistemology is being brought to the aid of an anti-subjectivist metaphysics, at just that point in the argument where the anti-subjectivism is moot. In any case, Churchland tends to *stress* knowing-how, while not *explicitly denying* the existence of knowing that - so that his position is unclear on the point. There is, however, as might be expected, a possible route from 'the epistemology that makes eliminative materialism possible' to a putative response to Nagel. Unlike the accumulation of putative cases of epistemic asymmetry in the absence of asymmetry of things known, discussed above, the argument from all knowledge being knowledge how is valid.[15]

Churchland now progresses his argument, making the claim that what is known via the two modes of knowing (the subjective and the objective) is identical - so that physicalism is redeemed, all entities and properties being open to investigation by science, and thus, ipso facto, being physical entities and properties. 'Our collective scientific enterprise', he states, while it won't detect the intrinsic character of the bat's (echolocationary) sensory states in the proprietary way in which the bat does (hence

the two modes), nonetheless 'will indeed both detect them (via microelectrodes) and represent them (in the language of science)'.[16] Once again, however, Churchland is begging the question: *if* science can achieve this outcome, *then* his position will be vindicated - there will be epistemic asymmetry in the absence of fundamental objective asymmetry. But whether science will, in fact, be able to perform this task is still an open question. If Nagel is right, then it is empirically impossible that science *will* ever achieve this outcome. So Churchland's conclusion at this point: 'the states represented, the bat's sensory states themselves, are presumably the same in each case ... [a]s before, the difference lies not in the character of the thing known; it lies in the distinct manner of the knowing'; is not supported by the argument from scientific speculation which he has advanced.[17]

Given that the title of Nagel's essay is 'What is it like to be a bat?', Churchland's claim here is very bold: in principle, science could tell us the answer to Nagel's question. There is no element of the bat's subjective awareness which is not capable of being represented objectively to us via science - so that science could, in principle, tell us what it is like to be a bat. [18]

Having considered Nagel's argument, Churchland now goes on to tackle Jackson's thought experiment, in which a neuroscientist called Mary's eyes:

> ... have high-tech chronic implants that flatten any spectral diversity in the incoming light. The only energy variations that get through to her retina are uniform across the entire spectrum ... she has therefore never seen the colour red as the rest of us have. She does not know what it is like to have a sensation of red. [19]

In consequence of having been in this position from birth, Mary cannot know, *via her auto-connected epistemic pathways*, what it is like to see red. Churchland will argue that this does not entail that she cannot know what it is like to see red (hence the significance of Mary's profession). His objection to Jackson's claim that she cannot know what it is like to see red will be based on the claim that once again epistemic asymmetry is being taken to accompany an asymmetry of things known.

Churchland sets out the situation in the now-familiar PDP terms:

> Mary's deprived condition has indeed kept her from ever knowing, via her auto-connected epistemic pathways, a sensation of red. No amount of neuroscientific book-learning on her part will ever constitute a representation-of-redness *in those pathways* ... ... if and when [she ever gains colour vision] ... and she is presented with a ripe tomato, then she will indeed come to know the sensation of red in a way that she has never known it before, in a way that finally exploits her autoconnected epistemic pathways'. (emphasis in original)[20]

Given the accuracy of the PDP account of cognition, this set of claims is uncontroversial. The emphasis on the phrase 'in those pathways' prepares the reader for what follows: as a neuroscientist, Mary has an alternative mode of knowing. Given that this mode of knowing can access all that is knowable via the mode of knowing which Mary lacks, she does in fact know what it is like to see red:

> Once more ... the genuine occurrence in Mary of that special, prescientific way of knowing does not entail that something non-physical is what gets known ... it is a kind of state with which she is scientifically quite familiar, a 70-20-30-Hertz coding triplet across the neurons of area V4, perhaps. That sensation is indeed new to her auto-connected sensory experience, but she's seen it a thousand times before in the

> auto-connected pathways of others. And it is the same thing in her as it is in all of the others: something physical.[21]

This reiterates a point which he has made earlier, apropos Leibniz: '... it remains possible that, should you and I happen to know what vectors constitute what sensations, and should we happen to know where and how to look for those activation vectors, then we might recognise those sensations ... as they go by. A well-informed observer could catch what an untrained observer could not'.[22] *If* Mary knows everything that there is to know about the physical process of colour coding, and *if* the experience of seeing colour is just this process which is exclusively physical (i.e. there's nothing over and above this), then the claim would appear to be supportable: Mary will recognise what the untrained observer would not - to the extent of even grasping what the untrained, and equally colour-blind, observer would not: an experience of what it is like to see red. This is ensured for Churchland by the fact that what Mary lacks is a mode of knowing, not access to an object of knowledge.

Robinson observes, regarding this strategy:

> the idea behind the 'mode' theory is, presumably, that if the phenomenal is merely another mode of access to respectable physical facts, then it does not itself constitute an extra fact: it is not an addition to ontology, but an extra mode of access to part of the previously accepted ontology.[23]

So, if we apply this to the 'Mary' case: employment of autoconnected epistemic pathways is merely another mode of access to the physical ontology to which she has access via another mode of knowing (i.e. her access to the autoconnected epistemic

pathways of *other* people). An 'ontology' here consists in a set of facts - so a 'mode of knowing' is not a fact. Robinson suggests that this must be rejected:

> Call the set of physical facts *P* and the set of scientific modes of access which [Mary] has to these facts *S*, and the ... mode she lacks *H*. If we regard *S* and *H* as external to *P*, then the addition of a new mode of access will not alter *P*'.[24]

Thus Churchland's claim must be that (given that a mode of knowing is not a fact) there's no addition to our ontology as a result of the acquisition of a new mode of knowing, in addition to the one already possessed. However, as Robinson goes on to point out, the physicalist thesis is that every fact is a physical fact - that is, all facts are already contained in *P*, where *P* is the set of all physical facts. But 'modes of access to physical facts are themselves simply physical processes and are included in physical facts'[25] (I take it that what Robinson means here is that it is a physical fact that that modes *S* and *H* are modes of access to *P* - so that this fact is ipso facto in *P*, which is, as just noted, the set of all facts, if physicalism is true. In other words, the opening claim that a mode of access to physical facts is not itself a physical fact 'is sophistical'. Robinson concludes:

> therefore, if [Mary] knows all the relevant members of *P* (i.e. knows all the facts that *P* comprises), then '[she] should know all the facts about *H*, including what the mode of access is phenomenally like.[26]

It's difficult to see how Churchland could respond satisfactorily to this objection. It would appear prima facie that he has two options. Given that his claim will be that there is *nothing additional* to know via *H* which Mary does not antecedently know via *S*, he could claim either that Mary can know what *H* is like vicariously, via her

knowledge of what *S* is like; alternatively, if *S* is not phenomenally like anything, then neither is *H*.

The first of these would require that we conflate (1) what it is like to see red, and (2) what it is like to see someone else's seeing red (where the latter is elliptical for 'what it is like to see the 70-20-30-Hertz coding triplet being activated in someone's neural area V4'). The claim that these experiences are identical could only be tested a posteriori, given sufficient development in the science of colour perception (and, perhaps, also given a suitable subject - a colour-blind neuroscientist who acquires the *H* mode - a capacity for colour perception via autoconnected epistemic pathways). Churchland is ignoring just this 'what is it like?' question, though (he doesn't, for example, express Mary's *S*-mode perception in terms of *what this is like* for Mary) - and in doing so, and concentrating rather on the 'what is known' question, he begs the question by assuming that what is know via *H* is nothing over and above what is known via *S*.[27]

The second of the options putatively open to Churchland - denying that either *S* or *H* *is* phenomenally like anything - would require that there are no phenomenal properties in our modes of knowledge. To use this response to Robinson's closing point that '[she] should know all the facts about H, including what the mode of access is phenomenally like' (op. cit.) would be to beg the question against Robinson. In any case, I can introspect - employ my auto-connected pathways - and thus confirm that there *is* something that it is like to see the colour red. So the second option must be

rejected.[28] I conclude that Churchland has failed to make his case for the absence of objects of knowledge which are in principle beyond the reach of physical science.

Churchland briefly considers the argument of a third opponent of his claim that the subjective will prove reducible to the objective: John Searle. In his *The Rediscovery of Mind*, Searle makes the important point that: 'any introspection I have of my own conscious state is itself that conscious state'.[29] On Churchland's construal, Searle thus identifies a fundamental difference between the two modes:

> ... one has direct and unmediated knowledge of the character of one's own sensations. In the case of physical things [Searle claims that] ... there is a legitimate distinction between appearance and reality.[30]

The latter part of this claim is one with which Churchland will agree wholeheartedly: his examples from the history of science are replete with instances of just this distinction operating in science, and the misleading appearance acting as an impediment to the re-partitioning of neuronal activation space which would yield a better theory. But the opening claim here - that one's knowledge of one's own sensations (via one's operation of the auto-connected pathways) yields knowledge which is 'direct and unmediated' cannot be accepted by Churchland. Churchland's model depends upon the two modes of knowing being distinct, but not fundamentally different. Specifically, the project of re-partitioning our neuronal activation space is the project of formulating a better theory of one's subjective states - so that the theory-ladenness of perception must extend to our subjective perception. On Searle's account, by contrast:

> ... in the case of the mental ... the appearance *is* the reality and vice-versa. One cannot be wrong about the contents of one's own mind.[31]

If this claim is accepted, then Churchland's PDP model is jeopardised (this in addition to the problems which it will create for his account of the reduction of the subjective to the objective). On the model, a 'theory' is a set of synaptic weights. Given that all activity within the network will be mediated by these weights, in addition to the pathways and prototypes also already established in the brain, the 'theory-ladenness of perception' consists in just this ubiquity of synaptic-weight determination of what is ultimately presented as output by the system. For me to access those objects of knowledge currently accessible only to Robinson's *H*-mode, and to do so in a way which did not involve this mediating influence of synaptic weight configurations (which is what the absence of the theory-ladenness of *H*-mode perception will require) is for me to perform a cognitive function which is beyond being accommodated by the model - and this failure to account for such 'direct and unmediated access' seems to be beyond accounting for in principle. If this is correct, then Churchland's need to reject Searle's claim is urgent.[32]

Churchland's response to Searle relies on the Freudian claim that: '... we are not infallible in our judgements about all mental states. One can misapprehend one's own desires and fears'.[33] Even if this conceded, this does not, however, serve Churchland's immediate purpose. 'Some' does not entail 'all': even if there is some empirical support for the claim that one's own apprehension of one's desires and fears is fallible - this in turn leading to the possibility of an *enriched* apprehension following

neural repartitioning (i.e. theoretical advancement), in just the way that Freud proposed, this does not entail that *all* of one's judgements regarding one's mental states are similarly infected by fallibility. So Churchland could be correct in his claim that 'the very concepts we bring to the business of apprehending our internal states are a source of chronic error', but without this pervading all such internal apprehension - so that Searle's claim is only partly erroneous, and Churchland's only partly correct.

What Churchland is ignoring by this stage in his exposition is the Nagel and Jackson question of the qualitative states which accompany one's sensations. One could surely concede the possibility of Churchland being correct with regard to every other introspectively-accessible object of knowledge, without having to concede on this particular object. Thus, I may be in error about my desires - but my awareness of what my occurrent desire *feels like* is surely infallible. By extension, my general awareness of what it is like to be me is a clear case of what Searle describes - namely, knowledge which is direct and unmediated.[34] Churchland has not yet dealt successfully with the subjective knowledge of what it is like to be that subject.[35]

If the conclusion - that 'what it is like to be me' is (perhaps uniquely) an object of knowledge of which my knowledge is direct and unmediated - is true, then this leaves a gap in the PDP account, as painstakingly constructed by Churchland. The truth of the claim will also entail that his claim that all knowledge is mediated is false. There is no further alternative of ignoring (while not denying) the fact that there is a fact of the matter concerning what it is like to have the sensation of being me. In his positive account of consciousness, however, Churchland effectively adopts this latter course,

by providing a putative set of features of consciousness which ignore the fact that there being something that it is like to be me is both a necessary and a sufficient condition for my being conscious.

In a passing aside, Churchland raises a technical quibble with the substance of Jackson's story:

> I will here ignore the developmental brain damage that would surely attend the chronic colour deprivation ascribed to Mary. Adulthood would be too late to set her free. By then her neuronal resources for colour vision would be severely atrophied. But this would spoil a good story.[36]

The situation here is more serious than this, however: if Churchland is correct in this assessment, then there is a problem for his position which is the counterpart of the *argumentum ad ignorantiam* weakness which he perceives to pervade his opponents' position. If Mary could not see red via her neuronal resources for colour vision following the removal of the implants, then she could not, in consequence, confirm that the two modes of knowing were accessing the same object of knowledge. But this then raises the question of who *could* do so. In the case of the non-impaired viewer, we would be unable to guarantee that what might be termed 'noise' - recurrent and/ or horizontal input from elsewhere in the brain, and deriving from activity associated with the *other* mode for accessing this object (Robinson's *H* mode) wasn't contributing the access to the object which is putatively being gleaned *solely* from the scientific, *S* mode. In the same way that the various parts of the brain cooperate with each other to minimise the cognitive loss arising from brain cell death or lesions in the brain, this process might, in the case in question, be both automatic, and unnoticed by the

cogniser herself.[37] In the case of the bat, we'd have nothing at all to compare the *S*-mode data *with* (for the obvious reason that we aren't bats). Confirmation of a precise 1:1 mapping of all that is accessed via *H* and all that is accessed via *S* may thus be impossible to achieve. Were my speculation about the parts of the brain confounding our endeavours by assisting each other to be false, and were Churchland's claim that Mary's colour neuronal resources would not be up to the task also to be false, then Mary could confirm the 1:1 mapping - and it might then be thought a safe presumptive inference to the more exciting claim that science could inform humans, via a suitably-trained up *S* mode, what it's like to be a bat. So - contra Churchland's proposal to ignore his own quibble rather than spoil a good story - my claim is that *his* story is the one which is spoiled.

If this is correct, then Churchland can never get beyond mere scientistic faith: when scientists eventually inform us that they know what it's like to be a bat, and that our assurance that this is true resides in an exact mapping of the bat's *H*-mode-discoverable objects of knowledge, and the scientists' *S*-mode-discoverable objects of knowledge, then we ought to accept the claim. In this case, the debate seems permanently intractable: one side will argue, on the basis of intuition, that the scientific claim is unreliable; the other side, that this is a mere *argumentum ad ignorantiam*.

This is where Churchland's claim that machine consciousness is an empirical possibility is useful. The question which we must ask is: is there anything that it would be like to be that machine? If this question is omitted - if Churchland merely alludes to the

claimed isomorphism between the machine's internal architecture, and that of human auto-connected pathways - then the burden of proof will remain with Churchland, who has not denied that one of my objects of knowledge is what it is like to be me. Nor has he denied that consciousness exists - and that what it is like to be me is knowledge (at present) exclusively accessible to the auto-connected pathways in my brain. All that he has denied is that these pathways provide the *only empirically possible* means of accessing this knowledge. The challenge for Churchland is, however, enormous: if he can support the claim that we can have a reliable means for confirming that there is something that it is like to be the machine, then the *H*-mode/ *S*-mode 1:1 mapping is vindicated for the machine. In the light of the machine's states being exclusively physical states (the condition for their comprehensive discovery via the *S*-mode), then the phenomenal property of there being something-that-it-is-like-to-be will have been demonstrated to be a physical property (at least for the machine). At that point, the claim that this property is a non-physical property in humans would be apparently impossible to sustain: dualism would be finally confuted. If this best-possible scenario for Churchland is not available, then the claim that all explanation is explanation of events which are open to explanation exclusively in terms of physical science and its explanatory categories is still contestable. The truth of this claim is, of course, a constraint on the truth of eliminative materialism per se: an autonomous mental category of explanation, underlying such processes as the formation of moral judgement, will destroy the whole project, with the exception of the claim that there are no sentences in the head. Having set out the task facing Churchland as he attempts to support the claim that machine consciousness is empirically possible, I now propose to evaluate his argument.

Churchland identifies 'the salient features of consciousness' as being:

i. possession of short-term memory;

ii. independence of sensory input (i.e. the brain continues to function in the absence of any new input from sources external to it);

iii. steerable attention (one can direct one's attention to a specific topic);

iv. capacity for alternative interpretations

v. the disappearance of consciousness in deep sleep

vi. its reappearance (albeit in muted or disjointed form) in dreaming; and

vii. the harbouring of the contents of the several basic sensory modalities within a single, unified experience.[38]

While any system which instantiates these features may be taken, in virtue of their instantiation, to be conscious, Churchland is careful to make clear that this set is not to be taken to be individually necessary and jointly sufficient for consciousness: it is, rather a 'rough triangulation'.[39]

Given this initial trepidation on Churchland's part, it is not entirely clear how to evaluate the claim here being made. What is conspicuous by its absence is what I have claimed to be the key issue: Churchland does not present the property of there being something that it is like to be the network in question as a condition for consciousness. I have suggested that this is necessary for consciousness (no truly conscious experience would be possible without an accompanying awareness of what this

experience is like - what it is like to be having this experience); I also take it to be sufficient for consciousness (if any individual has the experience of there being something that it is like to be that individual, then, for the duration of that experience, that individual must be conscious).[40]

This claim entails that the features in Churchland's list of factors are neither necessary nor sufficient for consciousness. I take it that one could suffer from some psychiatric condition where one's mind/brain did not instantiate any one of the features on Churchland's list, without thereby being unconscious (or incapable of instantiating consciousness) - so that none of the features is necessary.[41] Their sufficiency is more problematic: Churchland will go on to claim that all seven might be instantiated in a machine. But the 'steerability' of attention (feature (iii) in the list above) seems to be a condition which requires just the fundamental ontological subject/ object distinction which Churchland is concerned to deny. I have earlier objected to the presumed analogy between the mine detector and the human mind, on the grounds that the mine detector must be operated by an individual who is entirely separate from the machine itself (i.e. a sonar operator), but that no corresponding faculty for one part of the brain to access the other parts, and thus to act as a counterpart to the sonar operator is possible, or accounted for, in Churchland's account of the brain as a PDP network. [42] The only tenable solution is some form of ontological dualism - in which case there *is* the fundamental subject/ object distinction - and the subject can then instantiate the property of subjective self-awareness, and hence - on my account consciousness (this subjective self-awareness being the necessary and sufficient condition for consciousness). In the absence of this ontological distinction, Churchland

must give grounds for holding that this 'steerability' may be instantiated by a purely physical network.

The basis for the claim that this property - along with the others - may be so instantiated, is:

> ... a recent convergence in theoretical modelling and empirical brain research ... [which] suggests a way in which this might be done. A suitably configured recurrent network will display cognitive behaviours that are systematic functional analogues of all seven of these ... dimensions of conscious.[43]

Empirical brain research has apparently identified a massively recurrent network of axonal projections radiating from the brain's intralaminar nucleus. Churchland's account of how the features may be accounted for purely via the operation of this arrangement in the brain is disappointingly perfunctory. Steerability - our capacity to direct thought to a given topic, given the vast range of possibilities which the brain will accommodate - is accounted for in terms of the neural network's capacity:

> ... to enhance the chances of a specific recognition ... [by increasing] the probability that the appropriate prototype vector will be activated by the sensory inputs. Recurrent pathways can and do affect such activational probabilities by slightly pre-activating the relevant neuronal layer in the specific direction of some prototype vector ... (for example ... *choking sound* for the anxious mother).[44]

This would undoubtedly be a very useful feature - but the account given seems implausible, given that the initial decision is apparently attributable to the network itself, which elects to effect the requisite adjustment in probability. Churchland's account of this is obscure: '... attention is steerable by the network's own cognitive

activity', with no explanation of how this is initiated in the first place. In short, we have an example here of rational behaviour being left utterly mysterious: the model cannot account for the mother's deciding to pay particular attention to the sounds the baby is making, because she fears that the baby may choke. Such an account will reintroduce the queer entities and queer properties to which the traditional account will appeal. I conclude that *were* Churchland able to account for the steerability of attention, then he would have room in his account for the key property of subjective self-awareness: there is something that it is like to be a nervous mother - so that while the property of being nervous is occurrent, the mother is conscious. In the absence of these entities and properties, I cannot accept Churchland's claim that: 'a salient functional analogue *falls out of* the neurocomputational model ... for steerable attention'. (my emphasis) [45]

This claim is, in any case, heavily qualified later , when Churchland concedes that: 'I do not know, and you should not believe, that the preceding account is the correct account of consciousness. There is a remote chance, perhaps that it is'.[46] While Churchland is consistent in making this claim - that we can only speculate as to what is the correct account of consciousness - he is nonetheless confident that there is no category of the exclusively-subjective. This is clear from his claim that:

> ... [there is nothing] ... exclusively subjective about one's own mental states. Although they are typically known by way of one's auto-connected pathways, they can be known by way of other pathways as well. In fact, they are already so known, even by the standards of current common sense: other people infer my current mental state from my words, from my facial expression, and from my unfolding physical behaviour. The core point here is that there is simply no conflict

between being objective and being subjective. One and the same state can be both. [47]

Taking the opening claim here first - the denial of any exclusively subjective knowledge - it can be deduced from this that, even though we do not yet know what would constitute an accurate and comprehensive account of consciousness, such an account can in principle be furnished from a purely objective (physical, scientific) standpoint. This is a useful conclusion, as it allows Churchland to press his eliminative case despite the fact that he has to concede some difficulty in explaining what consciousness is. Whatever it is, it is to be accommodated by the explanatory categories of physical science. The immediate problem is that Churchland has no evidence to support this claim - which is based on nothing more than scientistic optimism. Even if he had succeeded in demonstrating that, for the various examples which he cited earlier, every object of knowledge accessible to the *H*-mode pathway was also accessible to the *S*-mode pathway, this would not entail that *anything* which is *H*-mode accessible is also *S*-mode accessible. But the induction which he here proposes has no basis at all: *none* of the cases cited (the configuration of one's limbs; the fullness of one's bladder and bowels; one's hair standing on end) *are* cases where the entire subjective experience is reducible to what is objectively accessible. In each case, the (ignored) condition of there being something that it is like to have the physical condition in question is beyond such objective knowledge. This is especially clear in the further claim made in the quotation above - that behaviouristic criteria of language and facial expression provide such objective access. To the very limited extent that this is true, the objective knowledge gained is unreliable and piecemeal - so that this cannot be held to support the claim that there is no fundamental dichotomy

between the subjectively-known and the objectively-known. Finally, Churchland's concluding sentence in this quotation: 'one and the same state can be both' is even weaker as a defence of the no-fundamental-dichotomy claim: this is consistent with the claim that there is no subjective knowledge which cannot be objective knowledge - but it does not entail this conclusion.

This weakness in Churchland's argument is significant - as the central claim of his next chapter ('Could an electronic machine be conscious?')[48] is that the answer is likely to be affirmative. But this optimism is based, not on a cogent account of what consciousness is, but rather on this unsubstantiated claim that, given that consciousness is a subjective condition, and given that there is no fundamental dichotomy between the subjective and the objective, so science can in principle both account for, and replicate, the requisite condition.

Churchland's preamble to the tentatively affirmative answer which he gives to the question incorporates a considerable amount of detail of artificial models which replicate the functioning of the human retina.[49] This leads Churchland to the optimistic speculation that such developments might be possible for the brain and nervous system as a whole. This then leads to the question: '... would such an artificial silicon brain, lodged in a robot body, be truly conscious?'. [50]

In considering the grounds for scepticism which have been raised by his opponents, Churchland reintroduces, as one of the grounds for scepticism: 'sensory qualia'. Having omitted these from his earlier account of consciousness, Churchland now notes

that: 'the problem is to find a plausible home for them within a purely physical framework'.[51]

Churchland here cites, with approval, the argument of Ned Block, who criticises functionalism in part on the basis of its failure to account for qualia.[52] Block's objection is that an abstract model which satisfies functionalist criteria for mentality, might meet all of the algorithmic requirements for responding to input stimuli, but that:

> ... sensory qualia would still be absent from this system. Functionalism, therefore, must be missing something important about the nature of conscious intelligence. [53]

Churchland endorses this claim of Block's: what is missing from the functionalist account is 'what takes place inside a cognitive creature'.[54] Once again, it seems, the sheer complexity of the brain's architecture holds the key to explaining a hitherto-perplexing feature of human mental existence:

> most of the major features of human and animal cognition arise not because of any program we are running. They arise because of the peculiar physical organisation of the nervous system, because of the way in which information is physically coded, and because of the physically distributed means by which that information gets transformed.[55]

Churchland is candid in his assessment that conscious robots are not in immediate prospect - but he nonetheless proposes that the answer to the question 'could an electronic machine be conscious?' is 'it rather looks that way'.[56] What has been overlooked is that Nagel and Jackson's claims regarding qualia have *not* been shown by Churchland to be false. Only by ignoring, at a key stage, the question of how we

know what it is like to be the cognizer that we are, and by making an unwarranted appeal to the likely success of science in accessing objectively all that is currently exclusively subjective, does he appear to refute the claim that qualia are *essentially* and *irreducibly* subjective. The reintroduction of qualia much later in the argument, coupled with the same scientistic optimism, cannot therefore be taken to sustain the claim that there will 'probably' be conscious machines.

Unless Churchland can present an argument for the objective accessibility of qualia, the entire eliminative materialism project is weakened - not just the immediate claim for the empirical possibility of artificial consciousness. Autonomous explanation, beyond the reach of science, will defeat the project of the elimination of the explanatory categories of folk psychology, or their reduction. That Churchland could salvage from this outcome the claim that there are no sentences in the head - this claim being consistent with the explanatory adequacy of folk psychology - merely serves to show how inadequate sentential eliminativism is as a form of eliminativism. My next chapter will finally address this question of what can be eliminated.

[1] William G. Lycan, *Mind and Cognition - A Reader*, Blackwell, Cambridge, Mass., 1990, p.441.
[2] Thomas Nagel, 'What Is It Like to Be a Bat?', in David M. Rosenthal (ed.): *The Nature of Mind*, Oxford University Press, New York, 1991, p.422.
[3] Paul M. Churchland, *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Mass., 1995, p.187.
[4] Patricia Smith Churchland, 'The Hornswoggle Problem' (paper given to The University of California San Diego Salk Institute 12.8.96). The use of the American term 'hornswoggle' in the title of this paper reveals a characteristically trenchant rejection of an opponent's position: Webster's Dictionary offers the synonym 'hoax' (*Webster's Third New International Dictionary*, 1986, p.1091.)
[5] Paul M. Churchland, *The Engine of Reason*, p.188.
[6] op. cit., p.191.
[7] Gottfried Wilhelm Leibniz, 'Monadology (1714)', in G.H.R. Parkinson (ed.), *Leibniz: Philosophical Writings*, Dent, London, 1973, p.181.
[8] Paul M. Churchland, *The Engine of Reason*, 1995, p.193.
[9] op. cit., p.196.
[10] op. cit., p.197.
[11] op. cit.
[12] op. cit.
[13] op. cit.
[14] Howard Robinson, 'The Anti-Materialist Strategy and the "Knowledge Argument"', in Howard Robinson (ed.), *Objections to Physicalism*, Clarendon, Oxford, 1993, p.164.
[15] This lack of clarity regarding Churchland's position extends to a lack of clarity regarding the related question of whether sentential or intentional eliminativism is to be favoured, as I suggested in an earlier chapter. With specific regard to the [knowledge that]/ [knowledge how] dichotomy, the problem would appear to be that [knowledge that] must have justification (i.e. any claim to the effect that '*(x)* knows that *(p)*' must entail that *(x)*'s commitment to the truth of *(p)* is justified: that *(x)* is 'in a position to know'). All of this appears irredeemably sentential, as already noted. But there is a further problem: the PDP network operates on patterns of purely physical stimulation, and its outputs are purely physical. But physical events seem intrinsically incapable of justifying anything. So either the physical events have semantic status, so that they can so justify, or [knowledge that] is not possible - so that it has to be eliminated, and all knowledge is [knowledge how]. If the first of these options is chosen by Churchland, then his eliminativism is, it seems, mere sentential eliminativism: when (appropriately) subjected to interpretation into sentential description, then (*qua* sentential particulars) the (translated) internal physical states of the system *do* serve to justify, so [knowledge that] is possible (as well as [knowledge how]). But this is a very modest outcome. If on the other hand the second option is chosen - the elimination of [knowledge that] - no obvious room is left in the account for subjectivity. As earlier noted, the criterion for evaluating [knowledge how] is entirely performative (*vide* Robinson's point that [knowledge that] is 'behaviouristic'). On this assessment, Churchland is faced with a dilemma: either settle for a hopelessly modest form of eliminative materialism - or opt for a strongly counter-intuitive radical form, where the subject is eliminated. A middle way would require [knowledge that] which could be justified by internal physical PDP

states which were intrinsically incapable of appropriate sentential description - but it's difficult to see how this might be achieved. If Churchland has noticed this (and he surely has - having discussed, for example, the status of 'moral facts'), then he is apparently remaining noncommittal, pending future empirical neurocomputational discovery.

[16] Paul M. Churchland, *The Engine of Reason,* p.199.

[17] op cit. As is often the case, Churchland appears to qualify his argument in an unexpected way here - it is only 'presumably' the case that the states are the same. Given that this is exactly the point that he needs - that the states are identical - this is an idiosyncratic use of the term 'presumably'. I assume that nothing hangs on it, as Churchland gives no reason for doubting that the states are identical.

[18] Churchland concludes his treatment of Nagel's bat with yet another clear piece of question-begging - 'if such non-physical features were to exist, why should one's autoconnected pathways pay any attention to them?. Those pathways are themselves entirely physical. How could they interact with any physical goings-on?' (op cit., p.200). *If* the PDP account - and physicalism generally - is true, then all cognition = physical causal processes in a physical medium. In this case, the mind-body problem in its conventional form would arise: how could non-physical phenomena have causal efficacy on this physical causal process? But that's exactly what is at issue here. The problem isn't merely the existence of 'queer' mental properties, epistemically accessible only to the individual whose queer properties they are: if there are such properties, and they are so accessible, then the PDP account is wrong - there are cognitive processes which cannot be accounted for by the model. Churchland cannot assume the truth of PDP in order to demonstrate the failure of Nagel's position: the truth of the PDP claim regarding the entirety of human cognition is still a hypothesis.

[19] op. cit., p.200.

[20] op. cit.

[21] op. cit., p.202.

[22] op. cit., p.193.

[23] Howard Robinson, 'The Anti-Materialist Strategy', p.165.

[24] op. cit.

[25] op. cit.

[26] op. cit., p.166 (I have amended Robinson's presentation here, as he unfortunately discusses a case other than the 'Mary' case which Churchland tackles - the 'deaf scientist'. The 'Mary' and 'deaf scientist' cases do not differ in any significant respect.

[27] I say that Churchland ignores the 'what is it like?' question, because he doesn't put the question of what it is like for Mary to have (*S*-mode) apprehension of *other people's* activation of the coding triplet. If he did do so, then to avoid begging the question against the existence of properties known via *H* which aren't known via *S*, he'd have to claim that Mary could know what it is like to see red (if there is anything that it's like) derivatively from knowing what it is like to see someone's 70-20-30 triplet being activated (were he to take my first option). It isn't surprising that Churchland doesn't explicitly avail himself of this option (i.e. the option of claiming that Mary knows what *H* is like derivatively from her knowledge of what *S* is like): the claim that there could be an exact correspondence between what it's like to see red,

and what it's like for a neuroscientist to witness the neural activity which is someone else's seeing red seems patently implausible.

[28] Churchland states, in 'The Direct Introspection of Brain States', in his *A Neurocomputational Perspective*, p.56: 'what interests me is the claim that reductions of various substances elsewhere in science *exclude the phenomenal features of the substance*. This is simply false ... redness, an objective phenomenal property of apples, is identical with a certain wavelength triplet of electromagnetic reflectance efficiencies'. So Churchland will not seek to avail himself of my option (2) - the elimination of the phenomenal *tout court*. (Though the reduction involved in identifying the phenomenal property of redness with a wavelength triplet - while it would grant Mary access to this property - seems once again to miss out what is at stake, namely the fact that there is something that it is like to have this neural event occur in one's brain, and that this is part of the phenomenal aspect of redness which is being identified with that neural event itself. The situation is thus ambiguous: Churchland claims not to eliminate the phenomenal - but I cannot see how his account can accommodate the phenomenal, so that he does appear, his protestations to the contrary notwithstanding, to eliminate the phenomenal.).

[29] John R. Searle, *The Rediscovery of Mind*, MIT Press, Cambridge Mass., 1992, p.97.

[30] Paul M. Churchland, *The Engine of Reason*, p.205. I take it that I have correctly identified the location for this claim in Searle: Churchland doesn't locate the claim in any specific text of Searle's. Although Searle doesn't seem to make Churchland's point explicit here, the claim that my introspection of a conscious state is that conscious state entails that there can be no appearance/ reality distinction on this occasion. In any case, what matters for my present purposes is Churchland's construal of a possible counter-position, rather than the questions of by whom the counter claim is made, and where.

[31] op. cit.

[32] Once again, the epistemology underpinning so much of eliminative materialism comes into play here: were there to be, as Searle claims, no appearance/ reality distinction for some class of mental phenomena, then Cartesian foundationalism might be in prospect - this de facto first person privilege being just what Descartes also claims. Certainty with regard to first-person knowledge would also rehabilitate truth, which Churchland has been concerned throughout to bracket (and would, I take it, render any attempt to eliminate [knowledge that] as a bona fide epistemic category hopeless). So, much rests on Churchland's ability successfully to rebut Searle's claim, beyond the immediate need to demonstrate the reducibility of the subjective to the objective.

[33] Paul M. Churchland, *The Engine of Reason*, p.205.

[34] Notwithstanding Churchland's distaste for a priori arguments which seek to delimit the possible future discoveries of science, it surely cannot be denied that my knowledge of what it is like to be me is *incorrigible*: any claim to the effect that 'he thinks that he knows what it is like to be him - but he's operating under a misapprehension, as will ultimately be demonstrated by new developments in neuroscience' is absurd.

[35] Though he does later - and illegitimately - claim that 'we have already defeated the ... arguments of Nagel and Jackson'. (op cit., p.250).

[36] Paul M. Churchland, *The Engine of Reason*, p.201.
[37] I have no evidence for this, but it seems to be in the spirit of Churchland's model.
[38] Paul M. Churchland, *The Engine of Reason*, p.213.
[39] op. cit. This cautious tone is later dropped, as I will go on to demonstrate.
[40] Churchland's earlier account, with its (speculatively-based) conclusion that science will detect and represent everything - every object of knowledge - available to the bat's unique *H*-mode, is consistent with knowledge of [what it is like to be a bat] being knowledge of some physical property. Churchland's ignoring the property of [what it is like to be a bat] seems to be based on the assumption that rebuttal of Nagel's claim that the property is accessible only to the bat (which Churchland has not in any case achieved) entails the elimination of the property itself. But this cannot be accepted. Churchland has in any case accepted (for example, in the Mary case), that the property of there being 'something that it is like' is a bona fide property. Churchland's claim against Nagel and Jackson is that epistemic asymmetry does not entail metaphysical asymmetry. But nor does the truth of the claim that everything accessible to the *H*-mode is accessible in principle to the *S*-mode entail that any property which Nagel has claimed to be accessible only to the *H*-mode does not exist. All that is entailed by the truth of this speculative claim is that *either the property is knowable via the S-mode, or it does not exist*. But this speculative claim does not justify the property's being ignored in what follows. Churchland must either make and defend the massively counter-intuitive claim that there is no such property, or, alternatively, make and defend the claim that the property *is* accessible to the *S*-mode, and thus physical.
[41] The research work of the psychologist Oliver Sacks, where patients suffered cognitive deficiencies apparently much more serious than those in Churchland's list - but were nonetheless conscious - would seem to bear out this intuition.
[42] Intriguingly, Churchland, in alluding to the 'self-connected way of knowing' that arises via the auto-connected pathways, giving rise to 'this peculiarly self-focused way of knowing about one's internal state', points out that 'it is part of any creature's internal system of body regulation ...' (*The Engine of* Reason, p.198). This point isn't further developed. The form of words here used is clearly dualistic - the creature regulating its body. But a non-dualistic account must leave something out: the auto-connected pathways are pathways via which *something* is in control of the body. Churchland surely evades the point with his question-begging claim that : 'such "auto-connected" ways of knowing have, as the *objects* of knowledge, exactly the same robustly physical things and circumstances as are occasionally known, through "heteroconnected" ways of knowing, by other individuals' (op. cit.). The 'way of knowing' at issue putatively *results* in an exclusively physical state, as he has argued. But as well as its consequences, the way of knowing will also have, among its objects of knowledge, the entities and properties which *initiated* the causal sequence which gives rise to this physical output. These mysterious entities and properties are swept up in the general claim that *all* objects of knowledge pertinent to this way of knowing are 'robustly physical'.
[43] op. cit., p.214.
[44] op. cit., p.218.
[45] op. cit.

[46] op. cit., p.223.
[47] op. cit., p.225.
[48] op. cit., pp. 227-252.
[49] op. cit., p.236. This work has been undertaken by Carver Mead, one of the original developers of integrated circuits.
[50] op. cit., p.243.
[51] op. cit., p.250. Churchland here makes the (false) claim that 'we have already defeated the negative arguments of Nagel and Jackson, so there is no need to readdress them'. The claim is thus that (a) qualia - as discussed by Nagel and Jackson - are really existing properties, but (b) they are not - contra Nagel and Jackson - exclusively subjective. This in turn entails that there is no principled reason why Mead's robot could not instantiate the property of having something that it is like to be that robot: what is accessible in the brain to scientific explanation is being taken by Churchland to be replicatable by science (this is, I take it, the point of the lengthy excursus on Mead's artificial retina) - so that the falsity of Nagel and Jackson's inaccessibility claim, together with the truth of the replicatability claim will entail the possibility of a robot with qualia. If there being something that it is like to be an object is, as I have earlier urged, sufficient for that object's being conscious, then the robot is in turn conscious.
[52] In his 'Troubles with Functionalism', in C.W. Savage (ed.), *Perception and Cognition - Issues in the Foundations of Psychology*, University of Minnesota Press, Minneapolis, 1978. This reinforces my earlier claim that Churchland's objection is not to the existence of qualia per se, but merely to the claim that they are objectively inaccessible.
[53] Paul M. Churchland, *The Engine of Reason*, p.251.
[54] op. cit.
[55] op. cit.
[56] op. cit., p.252.

CHAPTER 7 CONCLUSION: WHAT CAN BE ELIMINATED?

The 'radical synoptic claim' is the claim that the entirety of human mental existence can be accounted for in terms of the operation of activation vectors in human brains. This is entailed by the claims that the mind is the brain (the central physicalist claim) and that the brain's functioning is entirely in terms of PDP processing.[1] My last two chapters have argued for the falsity of the radical synoptic claim: Churchland's account of PDP cannot account for either moral agency or for consciousness. As the radical synoptic claim entails eliminative materialism, the falsity of the radical synoptic claim leaves the case for eliminative materialism at best not proven.

Eliminative materialists could respond to this by arguing that my rejection of a PDP account of moral agency or consciousness is based on an unwarranted aprioristic epistemology: they will deny that we can state a priori that PDP research will not eventuate in the solution of the apparent problems of moral agency and consciousness. As Rudder Baker points out, the scientific pragmatist holds that, for example:

> ... a science may set out to explain the impetus that keeps things in motion and end up abandoning the notion of impetus altogether.[2]

'The epistemology that makes eliminative materialism possible' thus accommodates Churchland's view of philosophy as proto-scientific speculation; his enthusiasm for PDP, in the light of its undeniable achievements to date in the case of artificial networks, together with his scientistic optimism, lead him to invest considerable optimism in the radical synoptic claim - this possibly to be vindicated by such future

developments as a conscious PDP machine.[3] Churchland will thus reject anything other than future empirical evidence as refuting the radical synoptic claim - so that its attendant entailment of eliminative materialism is also currently still intact.

In this chapter I will argue that there is no stable and coherent position for Churchland to occupy in the eliminativist debate - and that this is the source of his evident vacillation between a modest, conservative form of eliminativism, and a revolutionary form which promises a fundamental reappraisal of our species' self-conception. At the mildest end of a spectrum of eliminativist positions, PDP research findings support a position which may actually be endorsed by at least some Cartesian dualists. Churchland will himself wish to eliminate both this position, and the slightly more radical position which sees folk psychology as a systematically misleading theory which is a candidate for elimination - though this is perhaps not worth the effort, as folk psychology is no more inaccurate than is Newtonian mechanics (and may even be much less so). Once we come to positions which Churchland *can* endorse as being genuine full-strength eliminativism, the positions are so destructive of central tenets of our self-conception - and of philosophy - as to be incoherent.

Any eliminative materialist position must proceed from the sentential eliminativist claim that there are no 'sentences in the head'. The sentential eliminativist claim is derived from analogy with the mine detector. The architecture of its internal processes is incompatible with the syntax of natural language: for it to be compatible with natural language syntax, its processing would consist of sequences of inferences defined over propositions. As the instantiation of this syntax in a mechanical device would require

its manipulation of sentences, there are no surrogates for sentences inside the mine detector. Sentential eliminativism is thus established for the mine detector. Given that the PDP account is, as earlier argued, more psychologically realistic, the claim is then made that there are no sentences inside the heads of human cognizers either. In the same way in which the causally-efficacious states of the mine detector are not sentential entities, neither are the causally-efficacious states of the human brain.

Mere sentential eliminativism goes no further than to claim that there are no sentences in the head. This claim may well be true. But Churchland cannot endorse mere sentential eliminativism, as it will entail that, while there are no sentences in the head, it is nonetheless appropriate to utilise the intentional idiom to characterise what *is* going on in the head - so that intentional eliminativism is false. Churchland's commitment to theory-theory (the claim that folk psychology is a theory) will lead him to claim that it is the ontology of folk psychological theory which is being eliminated by sentential eliminativism - so that explanation in terms of the explanatory categories of folk psychology is misleading (as it entails realism about the ontology which underlies this type of explanation). The claim that folk psychology's explanatory categories ought not to be utilised - or ought only to be tolerated pending the development of some superior mode of explanation - is intentional eliminativism.

Mere sentential eliminativism thus fails to qualify as a form of eliminative materialism, because it has to accompany the falsity of the theory-theory - which, as I argue in my opening chapter, is necessary for eliminative materialism itself, as the latter commences from the claim that folk psychology is a candidate for elimination, and a necessary

condition for the truth of this claim is that folk psychology is a theory. So if mere sentential eliminativism is true, then the central thesis of eliminative materialism is false. There is, perhaps, an option of dropping the claim that folk psychology is a theory - but this would empty eliminative materialism of virtually all significance. If eliminative materialism amounts to no more than mere sentential eliminativism, it ought to be eliminated: the claim that there are no sentences in the head is endorsed by proponents of widely varying positions in philosophy of mind - including, *inter alia*, Cartesian dualists - so that the position is not materialist, and is only eliminative in a trivial sense.

Churchland must thus endorse intentional eliminativism, if the eliminative materialist thesis is to be sustained. Sentential eliminativism and theory-theory jointly entail intentional eliminativism.[4] As these two positions are individually necessary and jointly sufficient for intentional eliminativism, so intentional eliminativism entails the two positions. Unlike mere sentential eliminativism, intentional eliminativism is a *bona fide* form of eliminative materialism, as it thus entails the claims that the sentential descriptions of folk psychology constitute a theory, and that the theory is deficient, on account of the non-existence of its ontology.

As it stands, intentional eliminativism is a purely negative thesis. There are various strengths of intentional eliminativism: at the most modest extreme (what I will term 'merely pedantic intentional eliminativism'), the claim will be that *strictly speaking*, folk psychological categories ought not to be employed - or at least ought not to be employed by science, as they are less accurate than we would wish.[5] At the other

extreme (what I will term 'radical intentional eliminativism'), the claim is that folk psychological categories are fundamentally misleading, and must be expunged as soon as future scientific development permits. The intentional eliminativist position is presented in Churchland's seminal essay 'Eliminative Materialism and the Propositional Attitudes', which predates his work on PDP.:

> ... any declarative sentence to which a speaker would give confident assent is *merely a one-dimensional projection* - through the compound lens of Wernicke's and Broca's areas onto the idiosyncratic surface of the speaker's language - a one dimensional projection *of a four- or five-dimensional solid that is an element in his true kinematical state* ... being projections of that inner reality, such sentences do carry significant information ... and are thus fit to function as elements in a communication system.(emphasis mine)[6]

If we consider the issue in PDP terms, then the question at issue is that of the 'distance' in abstract vector space between folk psychological explanation, and a putatively superior explanation which might ultimately be available. If merely pedantic intentional eliminativism is hypothesised, then the distance is negligible. In this case, the distinction between sentential and intentional eliminativism is also negligible: both agree that there are no sentences in the head (intentional eliminativism entails sentential eliminativism); the latter will, unlike the former, incorporate the claim that folk psychology is a theory, but the prototypes proprietorial to the more favoured PDP theory will be 'close to' those of folk psychology.[7]

Merely pedantic intentional eliminativism is thus the mildest form of eliminativism, as it is the most modest position consistent with the eliminative materialist claims that folk psychology is a theory, and that it is a candidate for elimination on account of its being

a misleading theory, and hence potentially open to replacement by a superior theory. Churchland sets out what could be construed as being a merely pedantic position in 'Eliminative Materialism and the Propositional Attitudes, where he considers the possibility of a complete replacement of folk psychology:

> ... it is not inconceivable that some segment of the population ... should become intimately familiar with the vocabulary required to characterise our kinematical states, learn the laws governing their interactions and behavioural projections, acquire a facility in their first-person ascription, and displace the use of [folk psychology] altogether ...[8]

On this scenario, natural language, together with its syntactic forms and semantic categories, remains intact. In fact, this form of eliminativism is - consistent with its potentially merely pedantic status - remarkably conservative, as it will accommodate rationality, normativity, and the first-person perspective, all of which are alluded to in passing at this point in the essay.[9] All that *is* eliminated is the proprietary terminology and 'laws' of folk psychology. But, given that rationality is retained, the position must be conservative in further respects. To be rational - as this is conceived in folk psychological terms - is to formulate judgements and act in such a way as to maximise the consistency of one's actions and judgements with one's beliefs and desires. If we are to retain a conception of rationality, then it is difficult to see how we may deviate very far from this account. If we may continue to use natural language, but to allude to vectorial transformations as the basis for rational action, then these vectorial processes must be closely analogous to the deductive processes attributed by the folk psychologist to Pylyshyn's woman in the burning building. The merely pedantic formulation of intentional eliminativism would hold that entities *similar* to

beliefs and desires are operative in the rational decision of Pylyshyn's woman to flee the burning building (though actual description of her internal processes in these terms is misleading, *strictly speaking*). The similarity would reside in the fact that the actual states in the woman's brain encode content. Churchland could thus avoid the 'stimulus-responsish' problem: the woman's action is not a mere tropism, as she is responding to this content which is encoded in her brain. The encoding doesn't take the form of a mental sentence of the form parodically alluded to by Stich: '... a tiny television monitor or CRT [in the head] ... with thousands of English sentences displayed on the screen'.[10] But what *is* in the head may be, for instrumentalistic purposes, accurately described as being 'a belief that she is in imminent danger' - so that the fact that there is no sentence in her head of the form 'I am in imminent danger' does not serve to undermine the use of such natural language descriptions of internal cognitive states. The fact that we can retain our self-conception as rational beings (and hence as moral agents) is a clear strength of merely pedantic intentional eliminativism.

The fact that it is merely 'not inconceivable' that some segment of the population should master the seemingly enormous (and surely pointless) task of learning how to describe the neural processes and states which subserve the woman's action in deciding to flee, seems to underline the modesty of the eliminativism here on offer.

If what is *really* in the head is so close to the ontology of folk psychology (assuming that folk psychology *has* an ontology) as to accommodate rationality, then folk psychology cannot be seriously erroneous in its attribution of propositional attitudes in accounting for that rationality: the distance between folk psychology and the location

of what will become its superior replacement cannot be great, and the revolution is extremely limited. Merely pedantic intentional eliminativism is a form of eliminative materialism which Churchland would himself wish to eliminate. Given that theoretical replacement involves the discarding of some of the displaced theory (those elements which are not reduced by the successor), the question then is how much of folk psychology will survive. On the merely pedantic eliminativist account, a great deal would survive. For if it were to be endorsed, then folk psychology would, on this account, have a status comparable to that of Newtonian physics. While Einsteinian physics is a better theory, and Newton's is false in certain key respects (e.g. the relationship between space and time); Newton's theory is still taught in physics departments, and Newton is still revered as one of the precursors of an intellectual revolution which spawned, among other things, the modern social sciences. Indeed, it is consistent with Churchland's model that Einstein's theory awaits a similar fate. But a comparable status for folk psychology cannot be an outcome which Churchland will find congenial: his characterisation of folk psychology tends to be more in terms of medieval theories of witches and phlogiston - theories which were *fundamentally* inaccurate. *Radical* intentional eliminativism is, in fact, the motivation behind Churchland's entire eliminative materialist project:

> the real motive behind eliminative materialism is the worry that the "propositional" kinematics and "logical" dynamics of folk psychology constitute a radically *false* account of the cognitive activity of humans ... the worry is that our folk conception of how cognitive creatures represent the world (by propositional attitudes) and perform computations over these representations (by drawing sundry forms of inference from one to another) is a thoroughgoing *mis*representation of what really takes place inside us. (emphases in original)[11]

While Churchland might wish to see merely pedantic intentional eliminativism eliminated as a candidate position in theory of mind, non-eliminativists will likely regard the position as fatuous: probably no-one *will* claim that the propositional-attitude ascriptions of folk psychology are a literal and exhaustive account of what is going on in the head of Pylyshyn's woman. Jerry Fodor, whose 'language of thought' thesis is perhaps the most conspicuous target of (full-blooded) intentional eliminativism, presents what seems to be a rather tentative account of the description of internal cognitive reality which propositional attitude ascriptions present:

> *Claim 1* (the nature of propositional attitudes):
> For any organism *O*, and any attitude *A* toward the proposition *P*, there is a ('computational'/'functional') relation *R* and a mental representation *MP* such that
> *MP* means that *P*, and
> *O* has *A* iff *O* bears *R* to *MP*.[12]

Fodor commits himself to this position using appropriately cautious language: '[this theory is] ... ontologically committed to the attitudes and ... in my view [it] is *quite probably approximately true'*. (my emphasis)[13]

Fodor thus does not take propositional-attitude ascriptions to be literal and exhaustive as an account of what is going on in the head, and so may regard the merely pedantic form of intentional eliminativism with equanimity. The merely pedantic version of

intentional eliminativism may thus be safely ignored: it is far too conservative of folk psychology for Churchland to wish to endorse it, and it eliminates nothing which protagonists of folk psychology will likely wish to preserve. In short, as a possible theory of mind, it ought to be eliminated.[14]

While friends of folk psychology such as Fodor may regard merely pedantic versions of eliminative materialism with equanimity, Churchland's claim above that the motivation for the version of eliminative materialism which he will endorse is 'the worry that the "propositional" kinematics and "logical" dynamics of folk psychology constitute a radically *false* account of the cognitive activity of humans', will not be so regarded. Any version of eliminative materialism which retains rationality must thereby retain content: if the 'spontaneous' behaviour to which Churchland alludes is not based on PDP states in the brain which are representational states, then eliminative materialism entails a radical reassessment of our self-conception. This is the scenario which Fodor has in mind when he comments that:

> if commonsense intentional psychology were really to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species.[15]

There appears to be no stable position between merely pedantic intentional eliminativism and this catastrophic scenario. If rationality is to be retained, then at most eliminativists can claim that the falsity of folk psychology resides in its false claim that there are sentences in the head, and that what goes on in the head during cognition is computation over these sentences. The positive thesis of eliminative

materialism - if it is to retain rationality - must be that while there are no such sentential entities, and cognitive processing is thus not via sententially-based computation, the *actual* kinematics and dynamics of the brain preserve the truth of the claim that the woman flees the building because she smells smoke, perceives herself to be in imminent danger, and, in order to save herself, flees the building. But here the 'logical dynamics' of folk psychology are being preserved in the putatively more accurate account - so that eliminative materialism collapses into mere sentential eliminativism, combined with scientistic pedantry. If, on the other hand, the eliminativist's positive thesis is to avoid the 'merely pedantic' objection, then it must deliver on Churchland's promise of a radically different kinematics and dynamics. But if the dynamics of the new model are non-logical, then rationality (and thus freedom and morality) are ipso facto eliminated along with the propositional kinematics and logical dynamics intrinsic to folk psychology. Given that his central task is proto-scientific speculation, Churchland's position is inevitably subject to considerable uncertainty, but the absence of any stable position between the (eliminable) merely pedantic, and the (catastrophic) radical intentional eliminativism seems to answer what Churchland presents as an 'open question':

> 'we have here a novel paradigm ... whose kinematics and dynamics are radically *different* from that displayed in folk psychology. Whether it will prove capable of in some way reducing the familiar taxonomy of propositional attitudes is still an open question, but I am inclined to skepticism ... what this means is [that] ... we have a faltering old theory under siege by a new and more promising theory with an orthogonal categorial structure. That is why some of us anticipate the eventual elimination of our folk psychological ontology.[16]

It isn't clear what are the implications of the elimination of the ontology of folk psychology - but the 'catastrophic' fear seems initially borne out, when Churchland goes on to discuss 'truth' and 'reference' as possible candidate components of folk psychology's ontology. The extent to which a non-pedantic form of eliminativism is 'catastrophic' is a moot point. As we have already seen, Churchland, in his recent *The Engine of Reason, the Seat of the Soul*, is concerned to sustain a position which is simultaneously radical and conservative: a revolution in our self-conception is in prospect, and yet we will retain our freedom, capacity for moral choice, and consciousness. In response to the fact that many of its critics within the philosophical profession have regarded eliminative materialism as risible, precisely due to the perception that it proposes the catastrophic elimination of, *inter alia*, truth, belief, and consciousness, the Churchlands are similarly conservative in her essay 'Do We Propose to Eliminate Consciousness?':

> for the record, what eliminativist claims come to is essentially this: As science advances, certain "natural" categories that figured in an earlier theory turn out to have no role and no place in the replacing theory ... to put not too fine a point on it, the world is as it is; theories come and go, and the world remains the same. So theory modification does not entail that the nature of the world itself is modified *ipso facto*. It is our understanding of the nature of the world that gets modified.[17]

As it stands, it would be difficult to find fault with this commonsensical distinction between the elimination of some particular theory of consciousness, and the elimination of consciousness *per se*.[18] The key question is whether such a radical-yet-conservative option is sustainable. Similarly, Paul Churchland will seek to eliminate, not truth and reference *per se*, but rather the prototypes for truth and

reference which are proprietorial to folk psychology, and thus part of the ontology of this radically false theory. Thus Churchland confirms that:

> I am entirely willing to let go these notions, and to try to replace them with more penetrating evaluative/ semantical notions.[19]

The question must be whether it is coherent to argue for a 'successor concept' to truth which will both satisfy the radical aim of eliminating all of the detritus of the failed folk psychology, and yet simultaneously satisfy the conservative aim of retaining any of the characteristics of a concept which would be recognisably a concept of *truth*.[20] We cannot (currently) say in virtue of what this successor conception of truth *is* a conception of truth. Why, then, ought we to accept the possibility of such a development? When confronted by Putnam's objection that this is a mere 'gleam in Churchland's eye'[21], Churchland appears to miss Putnam's point:

> 'this is a serious mistake. A new kinematics and dynamics has been under vigorous exploration for roughly seven or eight years ...[22]

The availablity of an alternative account of the brain's kinematics and dynamics is consistent with merely pedantic intentional eliminativism: the conventional account of truth being retained, but truth being a relationship entered into by elements other than sentences, and preserved by a syntax other than computations over sentences. In missing the point of Putnam's objection, and shifting the focus of attention to progress in recent PDP research, Churchland appears to give the impression that discovery of a replacement concept is imminent.

There are no clear grounds for believing that there is a stable position between mere pedantry, and the catastrophic outcome raised by Fodor. When called upon to provide assurance that our conventional concepts of basic features of our mental lives such as consciousness and rationality are false, but that satisfactory replacement concepts which will preserve the underlying phenomena are a principled possibility, both Churchlands resort to mere affirmation that this is indeed the case - coupled with irrelevant allusion to the claimed progress to date achieved by PDP research. Even if we are to subscribe to the epistemology which sustains eliminative materialism, so that we reject the claim that it is analytic that rationality must involve content, and that, as content must be at least describable via natural language, so (at most) merely pedantic eliminative materialism is possible if there is to be the retention of rationality, it is still mere scientistic optimism which sustains the claim that there is a stable position between pedantry and catastrophe.

Once this is conceded, the scope for a sustainable form of eliminative materialism seems extremely tenuous. I have earlier argued that both mere sentential, and merely pedantic intentional, eliminativism ought both to be eliminated. The Churchlands themselves accept the case for elimination of catastrophic eliminativism, which would eliminate truth, rationality, morality, consciousness - and via these eliminations, philosophy (and the Churchlands themselves).

Once every other possible eliminativist position is itself eliminated on account of its being either non-eliminative (sentential eliminativism), too modest to be taken seriously (merely pedantic intentional eliminativism), or too extreme to be taken

seriously (catastrophic eliminativism), the remaining position - the precarious point somewhere midway between the pedantic and the catastrophic - is exposed for the implausible position that it is. In order to avoid being eliminated with the pedantic version, it must be radical (and so must eliminate significant parts of folk psychology's ontology); in order to avoid being eliminated with the catastrophic version, it must sustain recognisable concepts of truth, rationality, freedom, morality and consciousness. But the claim that some concept of, for example, rationality, *could* be too different from our procrustean, conventional notion of rationality to be captured, even inaccurately, by natural language and its employment of folk psychological categories - yet still be a replacement for rationality - is presented as an article of faith rather than as a demonstrable 'truth'. Once again, 'the epistemology that makes eliminative materialism possible' is pressed into service in defence of an implausible eliminativist claim: on this occasion, it does so by rejecting the analytic/ synthetic distinction on which the counter-intuition will rest. This scientistic faith is the inevitable recourse for Churchland, given his epistemological position, his elimination of the alternative eliminativist positions, and his insistence on the theory-theory. As Putnam - himself a defender of an epistemological position at least very similar to Churchland's - suggests:

> If I know that you know that the bus to work stops at the corner of Bartlett Avenue, and that you know that that is the nearest stop to your house, and that you dislike walking very far in the cold, I will naturally expect that if you decide to take the bus to work on a cold morning, you will wait for it at the corner of Bartlett Avenue.[23]

This is so clearly true - and so mundane in its simplicity - that it is surely beyond serious doubt. But - as Putnam goes on to insist:

> ... the failure to find a series of *scientifically describable events in the brain* ... which correspond point by point to the steps in [this] ... practical syllogism ... does not involv[e] a commitment to a categorical structure which is *incompatible with* ... that of brain science.[24]

It is the commitment to the claim that folk psychology is a theory which compels Churchland to insist, implausibly, upon the possibility of an alternative, non-propositional account of this sequence of rational behaviour which satisfies the seemingly impossible conditions of being both radical and conservative in the ways already set out in this chapter. Were he to renounce the theory-theory claim, then he could retain the claim that there are no sentences in the head, based upon PDP research to date, combined with the claim that the human brain is analogous to the mine detector, and sustain the conservative position that rationality and the other central features of our mental lives are beyond scientific disproof - by claiming that Putnam's account of the individual who waits for the bus is, as presented by Putnam, a psychological truism - but one which is subserved by a complex PDP process in the brain, rather than by computations over sentences inside the head. The (roughly accurate, and predictively adequate) natural language account presented by Putnam implies no commitment to actual sentences in the head, or their manipulation in thought.

What prevents this solution, and compels the untenable position which Churchland actually occupies is that such a position is not eliminativist: unless folk psychology is a

theory, it is not a candidate for elimination, and so eliminative materialism fails. As his commitment to theory-theory, and his need to steer a course between eliminativisms which are either eliminable on account of their weakness, or eliminable on account of their implausible radicalism, leads Churchland to this untenable position, I conclude that eliminative materialism ought itself to be eliminated.

[1] The radical synoptic claim could thus be argued to be false on one of two grounds: either that physicalism is false, and that a (possibly PDP) brain operates in conjunction with a non-physical mind; or that the brain is not an exclusively PDP system. This latter claim could acknowledge the plausibility of the claim that perception is via vectorial transformations from the sensory input level, but deny, for example, that, qua cognitive procedure, explanation is, as Churchland claims, an exclusively PDP procedure.

[2] Lynne Rudder Baker, 'Eliminativism and an Argument from Science' in *Mind and Language*, Vol. 8 No. 2 (Summer 1993), p.182.

[3] It is difficult to see what would count as evidence that a machine *was* conscious - but this is not of itself sufficient to refute the claim that a conscious machine is empirically possible.

[4] The claim that folk psychology is a theory entails the claim that folk psychology has a proprietory ontology: sentences in the head are the theoretical entities of folk psychology *qua* theory (this is a reconstruction of Churchland's position). It is thus the case that if the central claim of sentential eliminativism is true, and there *are no* sentences in the head, then the propositional attitude attributions of folk psychological practice allude to theoretical entities which do not exist (as did medieval witch theories). At the very least there is thus scope for being misleading in employing such attributions, so that at the very least there is a case for what I refer to in this chapter as 'merely pedantic intentional eliminativism'.

[5] In the absence of a PDP-based alternative mode of description for states currently characterised as, for example, 'believing that Churchland has a good case', then we have no alternative but to state that such folk psychological formulations are 'less accurate than we would wish'.

[6] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', in his *A Neurocomputational Perspective - The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Mass., 1989, p.18. (This essay originally published in the *Journal of Philosophy* 78 (1981); reprinted as chapter 1 of his 1989 collection. Page references in this chapter are to the latter.).

[7] It must be conceded, however, that this seems implausible, given that the folk psychological formulation is a 'one dimensional projection' of what is in fact a 'four- or five-dimensional solid'. Churchland's inability to embrace merely pedantic eliminativism is further underlined by the fact that in the *Neurocomputational Perspective* version of the essay, he usefully adds 1989 comments throughout the by then eight-year old original text. At this point in the exposition Churchland adds that 'this guess (i.e. the guess that the relationship would be of a one-dimensional projection of a four- or five-dimensional solid) ... has proved to be very timid. The relevant cognitive statespaces typically have hundreds, thousands, or even millions of distinct dimensions, and their partitioning into hypersolids is correspondingly complex'. ('Eliminative Materialism and the Propositional Attitudes', p.17). Churchland's commitment to the radical synoptic claim for PDP further commits him to a correspondingly radical intentional eliminativism.

[8] Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', p. 18.

[9] Churchland suggests that we must 'transcend the poverty of [folk psychology's] conception of rationality by transcending its propositional kinematics

entirely' (op. cit., p.16); and that folk psychology's declarative sentences ... 'are unfit to represent the deeper reality in all its kinematically, dynamically, *and even normatively relevant* aspects' (my emphasis) (p.18).

[10] Stephen P. Stich, *From Folk Psychology to Cognitive Science - The Case Against Belief*, MIT Press, Cambridge, Mass., 1983, p.36.

[11] Paul M. Churchland, 'Activation Vectors versus Propositional Attitudes: How the Brain Represents Reality', *Philosophy and Phenomenological Research*, Vol. LII, No. 2, June 1992, p.420.

[12] Jerry A. Fodor, *Psychosemantics - The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge, Massachusetts, 1987, p.17

[13] Jerry A. Fodor, *Psychosemantics*, p.16.

[14] It may seem surprising that Fodor can, on my account, be sanguine about merely pedantic intentional eliminativism, given that this position is - unlike mere sentential eliminativism - consistent with the central claim of eliminative materialism (a claim which Fodor, as a theory dualist, must reject). In his essay 'Varieties of Eliminativism' (in *Mind and Language*, Vol.8, No.2), Andy Clark suggests that Fodor's language of thought is what is eliminated by mere sentential eliminativism. Fodor's position is that the causal processing in the brain of the physical states which are the mental representations, will preserve the syntactical relationships of natural language. As the mental representations preserve the semantics of natural language, there is thus a de facto 'language of thought'. This is a much more elaborate claim than Stich's tiny television screens. But Fodor can retain his language of thought, and hence reject a version of mere sentential eliminativism which proposes the elimination of natural-language-syntax-in-the-head, while nonetheless accepting the extremely weak merely pedantic intentional eliminativist claim that propositional attitude ascriptions don't convey the full story of what is going on while this syntactic procedure is actually operative.

[15] Jerry A. Fodor, *Psychosemantics*, p.xii.

[16] Paul M. Churchland, 'Activation Vectors versus Propositional Attitudes', p.421.

[17] Paul M. Churchland and Patricia Smith Churchland, 'Do We Propose to Eliminate Consciousness?', in Robert N. McCauley (ed.) *The Churchlands and their Critics*, Blackwell, Cambridge Massachusetts, 1996, p.297.

[18] Rather disappointingly, the Churchlands do not avail themselves of the opportunity here to provide an alternative account of consciousness. The suspicion remains that, as the endorsement of radical eliminativism must entail the elimination of the irreducibly subjective, so the claim that consciousness is not to be eliminated must either entail the falsity of radical eliminativism, or some as-yet undiscovered account of consciousness which is fundamentally distinct from our conventional account, but which nonetheless preserves *consciousness*.

[19] Paul M. Churchland, 'Activation Vectors versus Propositional Attitudes', p.422. This appears to be a clear case of what Andy Clark has referred to as Churchland's 'dealing in futures' (Andy Clark, 'Dealing in Futures: Folk Psychology and the Role of Representations in Cognitive Science', in Robert N. McCauley (ed.), *The Churchlands and their Critics* (Blackwell, Cambridge Mass., 1996). My objection is that Churchland need not make the attempt to find 'more penetrating' semantical notions. To be semantical is to be characterisable in terms of content, so that any such

successor notion would be incompatible with intentional eliminativism - and hence incompatible with eliminative materialism. As with qualia, freedom, and normativity, the claim that what I have referred to as a 'radical -yet-conservative' option is sustainable in principle is false - and this falsity is, *pace* Churchland's post-Quinean epistemological stance, demonstrable a priori.

[20] The same claim stands for rationality, consciousness, and all other aspects of our mental lives currently infected by their conception via the categories of folk psychology.

[21] Hilary Putnam, *Representation and Reality*, MIT Press, Cambridge Mass., 1988, p.110.

[22] Paul M. Churchland, 'Activation Vectors versus Propositional Attitudes', p.420.

[23] Hilary Putnam, 'Truth, Activation Vectors and Possession Conditions for Concepts', in *Philosophy and Phenomenological Research*, Vol. LII, No. 2, June 1992 , p.439.

[24] op. cit., p.440.

BIBLIOGRAPHY

Lynne Rudder Baker: 'Eliminativism and an Argument from Science', in *Mind and Language* vol. 8, no. 2, Summer 1993.

Ned Block: 'Troubles with Functionalism', in C.W. Savage (ed.): *Perception and Cognition - Issues in the Foundations of Psychology*, University of Minnesota Press, Minneapolis, 1978.

David Charles and Kathleen Lennon (eds.): *Reduction, Explanation, and Realism*, Clarendon Press, Oxford, 1992.

William Charlton, *The Analytic Ambition*, Blackwell, Oxford, 1991.

Patricia Smith Churchland, 'Epistemology in the Age of Neuroscience', in *The Journal of Philosophy*, 1987.

Patricia Smith Churchland, 'The Hornswoggle Problem' (paper given to The University of California San Diego Salk Institute 12.8.96).

Patricia Smith Churchland: *Neurophilosophy*, MIT Press, Cambridge, Mass., 1986.

Patricia Smith Churchland and Terrence J. Sejnowski: 'Neural Representation and Neural Computation' in William G. Lycan (ed.): *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Massachusetts, 1990.

Paul M. Churchland: 'Activation Vectors versus Propositional Attitudes: How the Brain Represents Reality', *Philosophy and Phenomenological Research*, Vol. LII, No. 2, June 1992.

Paul M. Churchland: 'The Continuity of Philosophy and the Sciences', *Mind and* Language, Vol. 1, No. 1, 1986.

Paul M. Churchland: *A Neurocomputational Perspective - The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Mass., 1989.

Paul M. Churchland: 'The Direct Introspection of Brain States', in his *A Neurocomputational Perspective*.

Paul M. Churchland: 'Eliminative Materialism and the Propositional Attitudes, in William G. Lycan (ed.): *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Massachusetts, 1990.

Paul M. Churchland: *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Mass., 1995.

Paul M. Churchland: 'Evaluating our Self-Conception', in *Mind & Language* vol. 8, no. 2, Summer 1993.

Paul M. Churchland: 'Explanation: A PDP Approach' in his *A Neurocomputational Perspective*.

Paul M. Churchland: 'Folk Psychology' in Samuel Guttenplan (ed.): *A Companion to the Philosophy of Mind*, Blackwell, Oxford, 1994.

Paul M. Churchland: 'Learning and Conceptual Change', in his *A Neurocomputational Perspective*.

Paul M. Churchland: *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*, (Revised edition), MIT Press, Cambridge Massachusetts, 1988.

Paul M. Churchland: 'Moral Facts and Moral Knowledge', in his *A Neurocomputational Perspective*.

Paul M. Churchland: 'On the Nature of Theories: A Neurocomputational Perspective', in his
A Neurocomputational Perspective.

Paul M. Churchland: 'Postscript: Evaluating Our Self-Conception', appended to his 'Eliminative Materialism and the Propositional Attitudes' in Paul K. Moser and J. D. Trout (eds.): *Contenporary Materialism - A Reader*, Routledge, London, 1995.

Paul M. Churchland: *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge, 1979.

Paul M. Churchland and Patricia Smith Churchland: 'Clark's Connectionist Defense of Folk Psychology', in Robert N. McCauley (ed.): *The Churchlands and their Critics*.

Paul M. Churchland and Patricia Smith Churchland: 'Do We Propose to Eliminate Consciousness?', in Robert N. McCauley (ed.): *The Churchlands and their Critics*.

Paul M. Churchland and Patricia Smith Churchland: 'Flanagan on Moral Knowledge' in Robert N. McCauley (ed.): *The Churchlands and their Critics*.

Paul M. Churchland and Patricia Smith Churchland: 'Intertheoretic Reduction: a Neuroscientist's Field Guide', in Richard Warner and Tadeusz Szubka (eds.): *The Mind-Body Problem - A Guide to the Current Debate*, Blackwell, Oxford, 1994.

Andy Clark, 'Dealing in Futures: Folk Psychology and the Role of Representations in Cognitive Science', in Robert N. McCauley (ed.): *The Churchlands and their Critics* (Blackwell, Cambridge Mass., 1996.

Andy Clark, 'The Varieties of Eliminativism: Sentential, Intentional and Catastrophic', in *Mind & Language* vol. 8, no. 2, Summer 1993.

Adrian Cussins: 'Nonconceptual Content and the Elimination of Misconceived Composites!', in *Mind and Language* vol. 8, no. 2, Summer 1993.

Jerry A. Fodor: 'The Big Idea: Can There be a Science of Mind?', in *The Times Literary Supplement*, 3.7.92.

Jerry A. Fodor: 'Fodor's Guide to Mental Representation' in his *A Theory of Content and Other Essays*, MIT Press, Cambridge Mass., 1990.

Jerry A. Fodor: 'Observation Reconsidered', in *A Theory of Content and Other Essays*, MIT Press, Cambridge, Mass., 1990

Jerry A. Fodor: *Psychosemantics - The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge, Mass., 1987.

John D. Greenwood (ed.), *The Future of Folk Psychology*, Cambridge University Press, Cambridge, Mass., 1991.

Barbara Hannan: 'Don't Stop Believing: The Case Against Eliminative Materialism', in 'Don't Stop Believing: The Case Against Eliminative Materialism' in *Mind & Language* vol. 8, no. 2, Summer 1993.

Ted Honderich (ed.): *The Oxford Companion to Philosophy,* Oxford University Press, Oxford, 1995.

Terence Horgan and John Tienson: *Connectionism and the Philosophy of Psychology,* MIT Press, Cambridge, Mass., 1996.

Thomas S. Kuhn: *The Structure of Scientific Revolutions*, (second edition), University of Chicago Press, Chicago, 1970.

Gottfried Wilhelm Leibniz: 'Monadology (1714)', in G.H.R. Parkinson (ed.): *Leibniz: Philosophical Writings*, Dent, London, 1973.

William G. Lycan: *Mind and Cognition - A Reader*, Basil Blackwell, Cambridge, Mass., 1990.

Robert McCauley: 'Explanatory Pluralism and the Co-evolution of Theories in Science', in his *The Churchlands and their Critics*, Blackwell, Oxford, 1996.

Geoffrey Madell: *Mind and Materialism*, Edinburgh University Press, Edinburgh, 1988.

Thomas Nagel: 'What Is It Like to Be a Bat?', in David M. Rosenthal (ed.): *The Nature of Mind*, Oxford University Press, New York, 1991.

A. Partington (ed.): *The Oxford Dictionary of Quotations*, Revised 4th edition, Oxford University Press, Oxford, 1996.

Hilary Putnam: *Reason Truth and History* Cambridge University Press, Cambridge, 1981.

Hilary Putnam: *Representation and Reality*, MIT Press, Cambridge Mass., 1988.

Hilary Putnam: 'Truth, Activation Vectors and Possession Conditions for Concepts', in *Philosophy and Phenomenological Research*, Vol. LII, No. 2, June 1992.

Zenon Pylyshyn: 'Cognitive Representation and the Process-Architecture Distinction', in *Behavioural and Brain Sciences* 3, 1980.

W.V.O. Quine: 'Epistemology Naturalized' in Hilary Kornblith (ed.): *Naturalizing Epistemology*, MIT Press, Cambridge, Mass., 1994.

Ramsey, W., Stich, S., and Garon, J., 'Connectionism, Eliminativism, and the Future of Folk Psychology' in John D. Greenwood (ed.), *The Future of Folk Psychology*, Cambridge University Press, Cambridge, Mass., 1991

Howard Robinson, 'The Anti-Materialist Strategy and the "Knowledge Argument"', in Howard Robinson (ed.): *Objections to Physicalism*, Clarendon, Oxford, 1993.

David M. Rosenthal (ed.): *The Nature of Mind*, Oxford University Press, New York, 1991.

Wesley C. Salmon: '[The] Epistemology of Natural Science' in Jonathan Dancy and Ernest Sosa (eds.): *A Companion to Epistemology*, Blackwell, Oxford, 1992.

Roger Scruton: *Modern Philosophy - A Survey*, Sinclair-Stevenson, London, 1994.

John R. Searle: *The Rediscovery of Mind*, MIT Press, Cambridge, Mass., 1992.

Wilfrid Sellars: 'Empiricism and the Philosophy of Mind', in his *Science, Perception and Reality*, Routledge, London, 1963.

Michael Smith: 'Realism', in Peter Singer (ed.): *A Companion to Ethics,* Blackwell, London, 1991.

Tom Sorell: *Scientism - Philosophy and the Infatuation with Science*, Routledge, London, 1991.

Stephen P. Stich: *From Folk Psychology to Cognitive Science - The Case Against Belief*, MIT Press, Cambridge, Mass., 1983.

Stephen Stich: 'What is a Theory of Mental Representation?', in Stephen Stich and Ted A. Warfield (eds.): *Mental Representation - A Reader*, Blackwell, Oxford, 1994.

Richard Warner and Tadeusz Szubka (eds.), *The Mind-Body Problem - A Guide to the Current Debate*, Blackwell, Oxford, 1994.

Webster's Third New International Dictionary, Merriam Webster, Chicago, 1986.

Ludwig Wittgenstein: *Tractatus Logico-Philosophicus*, D.F. Pears and B.F. McGuinness (eds.), Routledge, London, 1961.